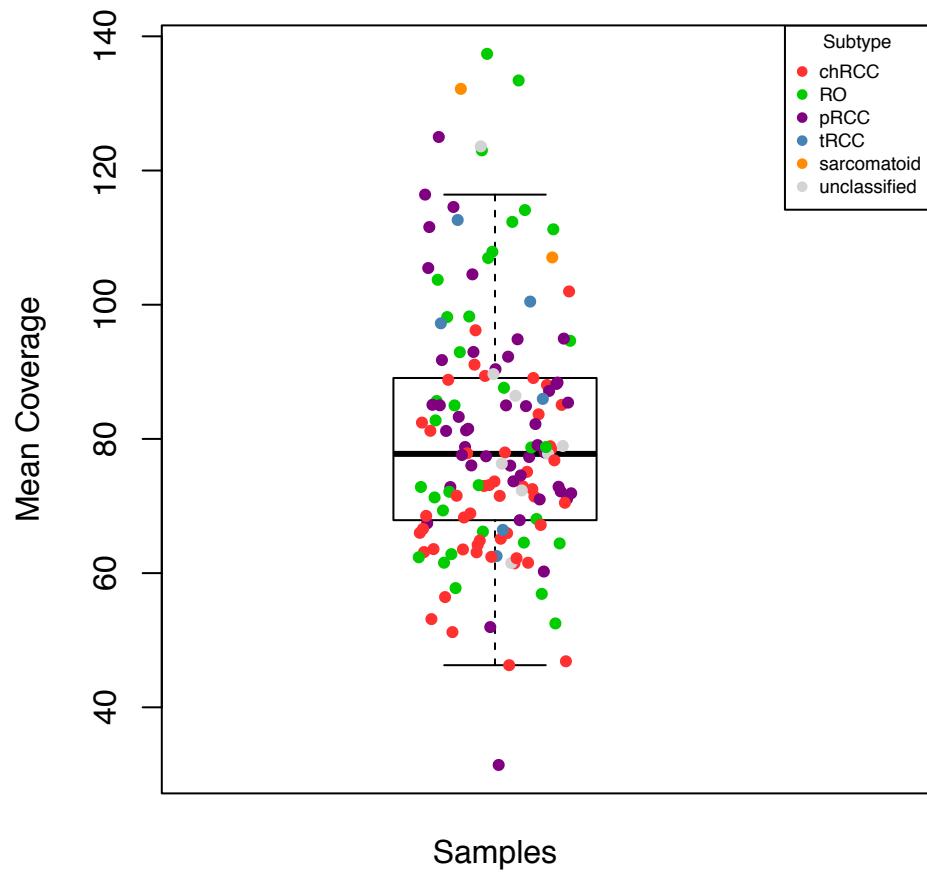
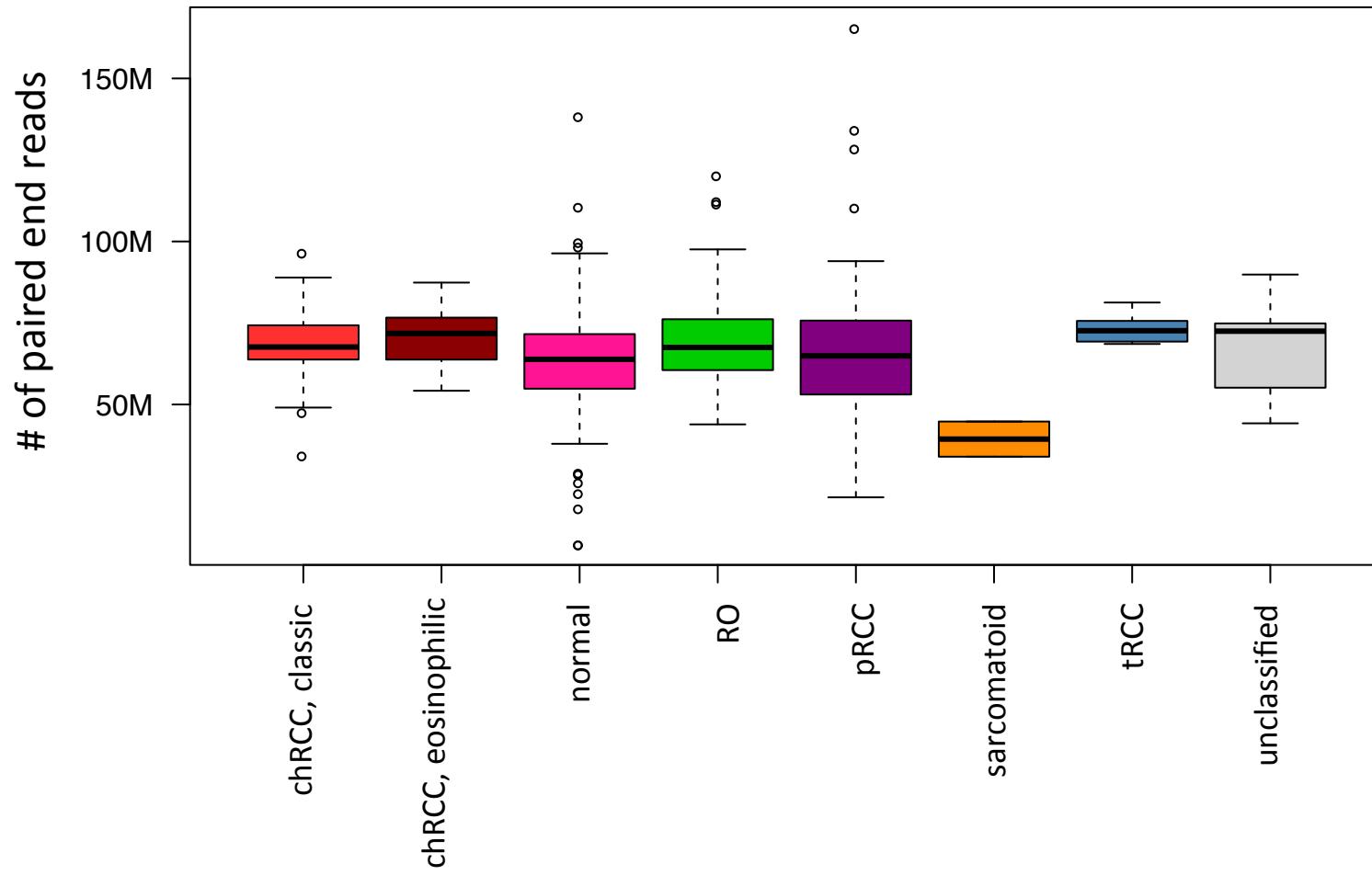


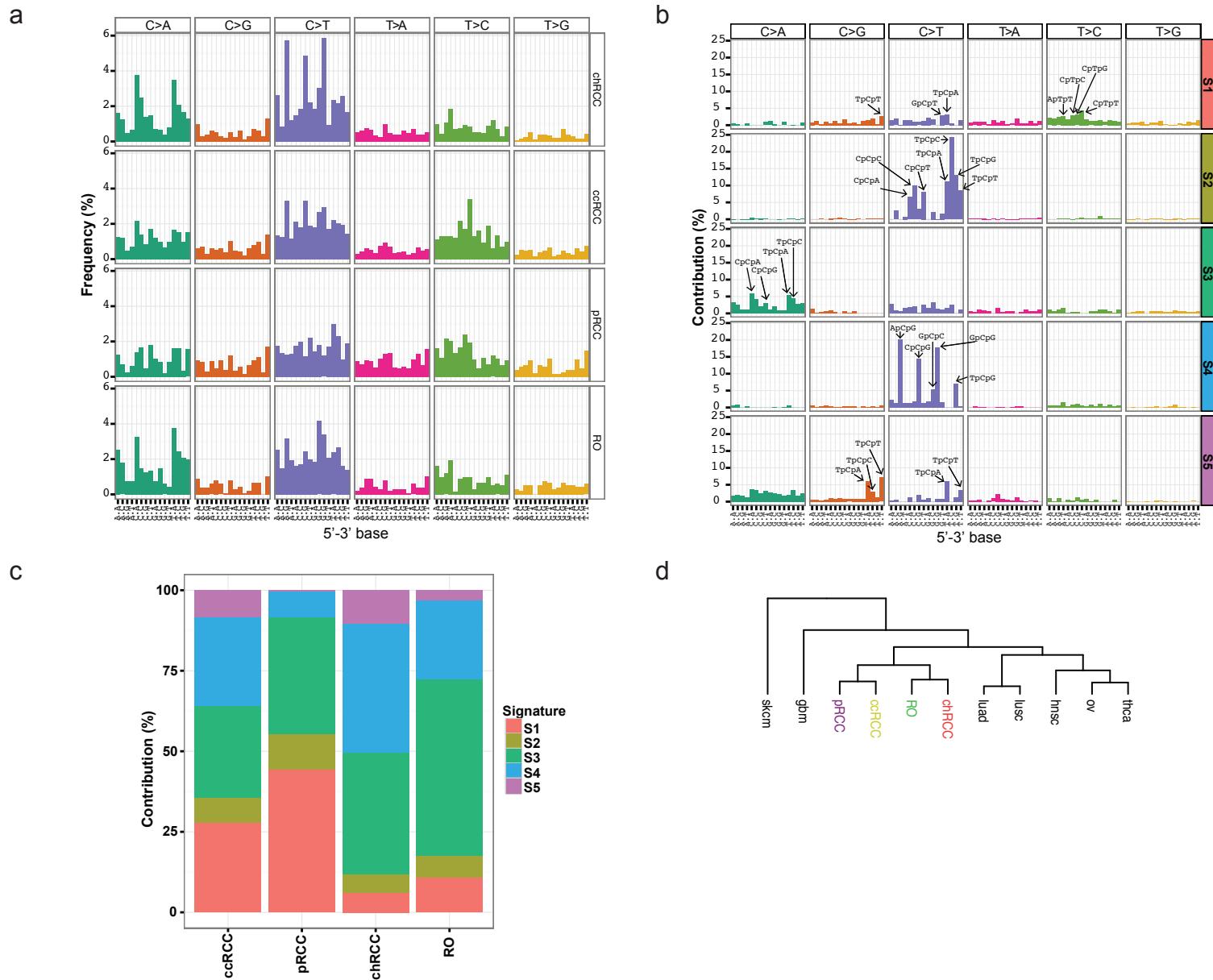
**Supplementary Fig. 1.** Graphical representation of the data types obtained by sample.



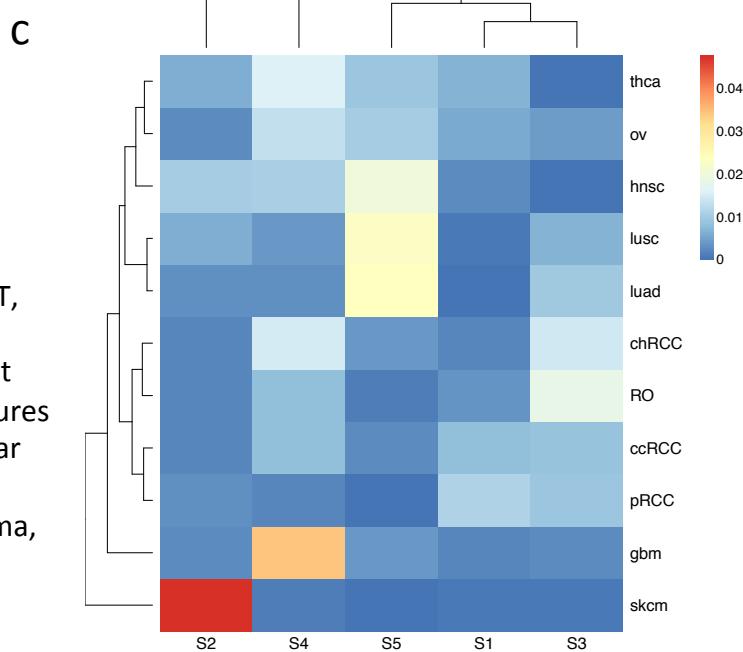
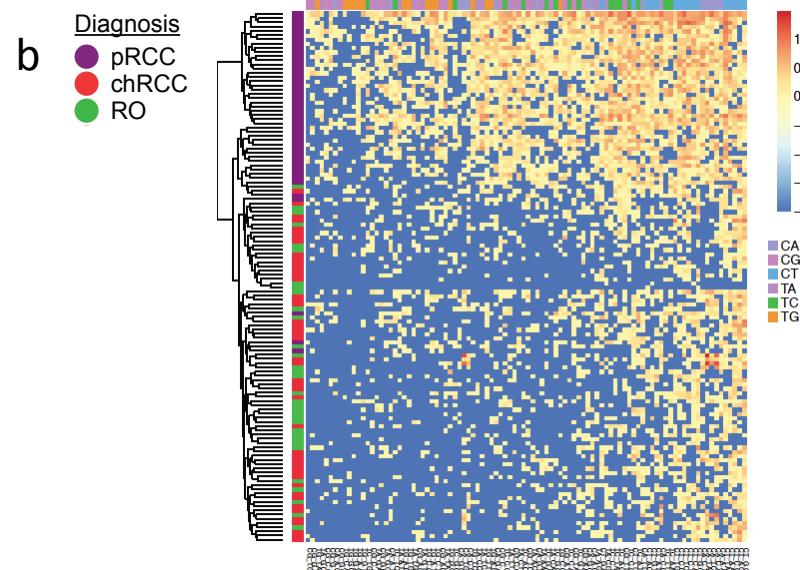
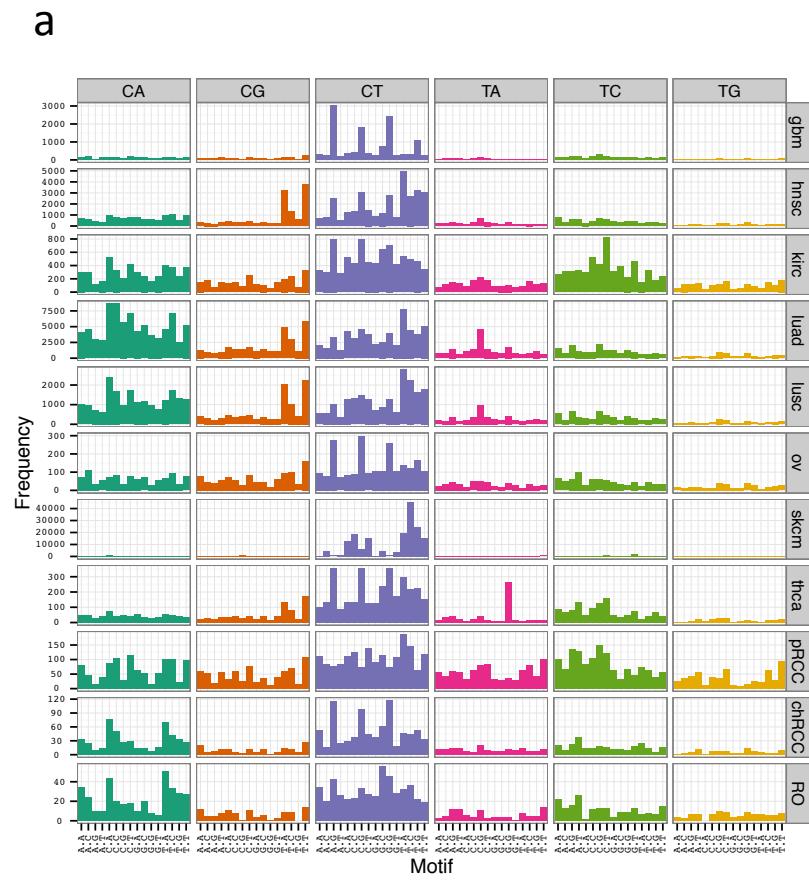
**Supplementary Fig. 2.** Boxplot depicting the exome coverage across samples. Boxplot defined in **Fig.1** legend in the main text.



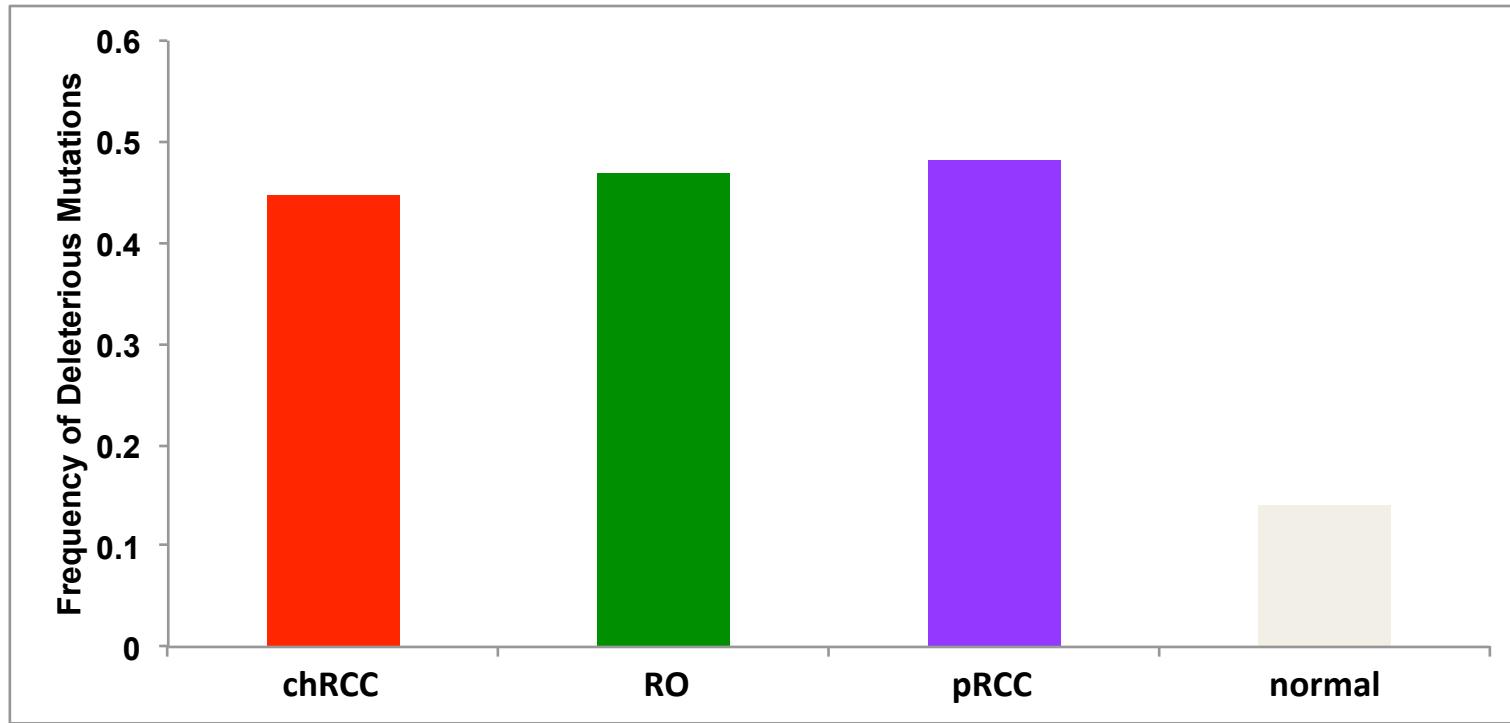
**Supplementary Fig. 3.** Boxplot depicting the RNA-seq depth across the samples. Average of 68 million paired end reads were obtained for the tumor samples. Boxplot defined in **Fig.1** legend in the main text.



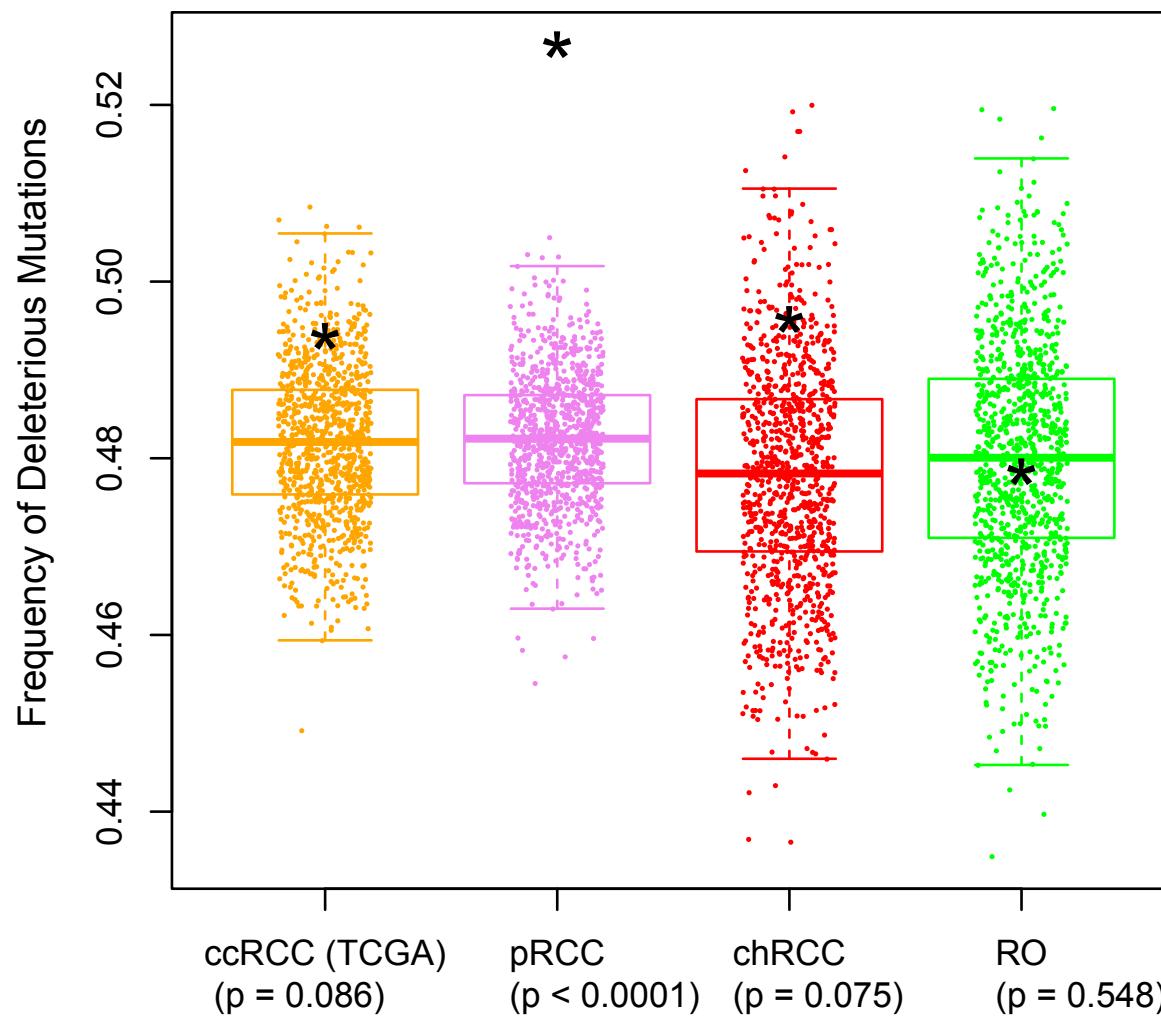
**Supplementary Fig. 4.** (a) Frequency of each mutation-type (C>A, C>G, C>T, T>A, T>C, T>G) in the context of 5' and 3' flanking bases. (b) Five mutation signatures identified by NMF (c) Contribution of the mutations signatures in RCC (d) Hierarchical clustering of cancer type based on mutational signatures (thca – thyroid carcinoma, ov - ovarian serous cystadenocarcinoma, hnsc – head and neck squamous cell carcinoma, lusc – lung squamous cell carcinoma, lujad – lung adenocarcinoma, gbm - glioblastoma multiforme, skcm – skin cutaneous melanoma).



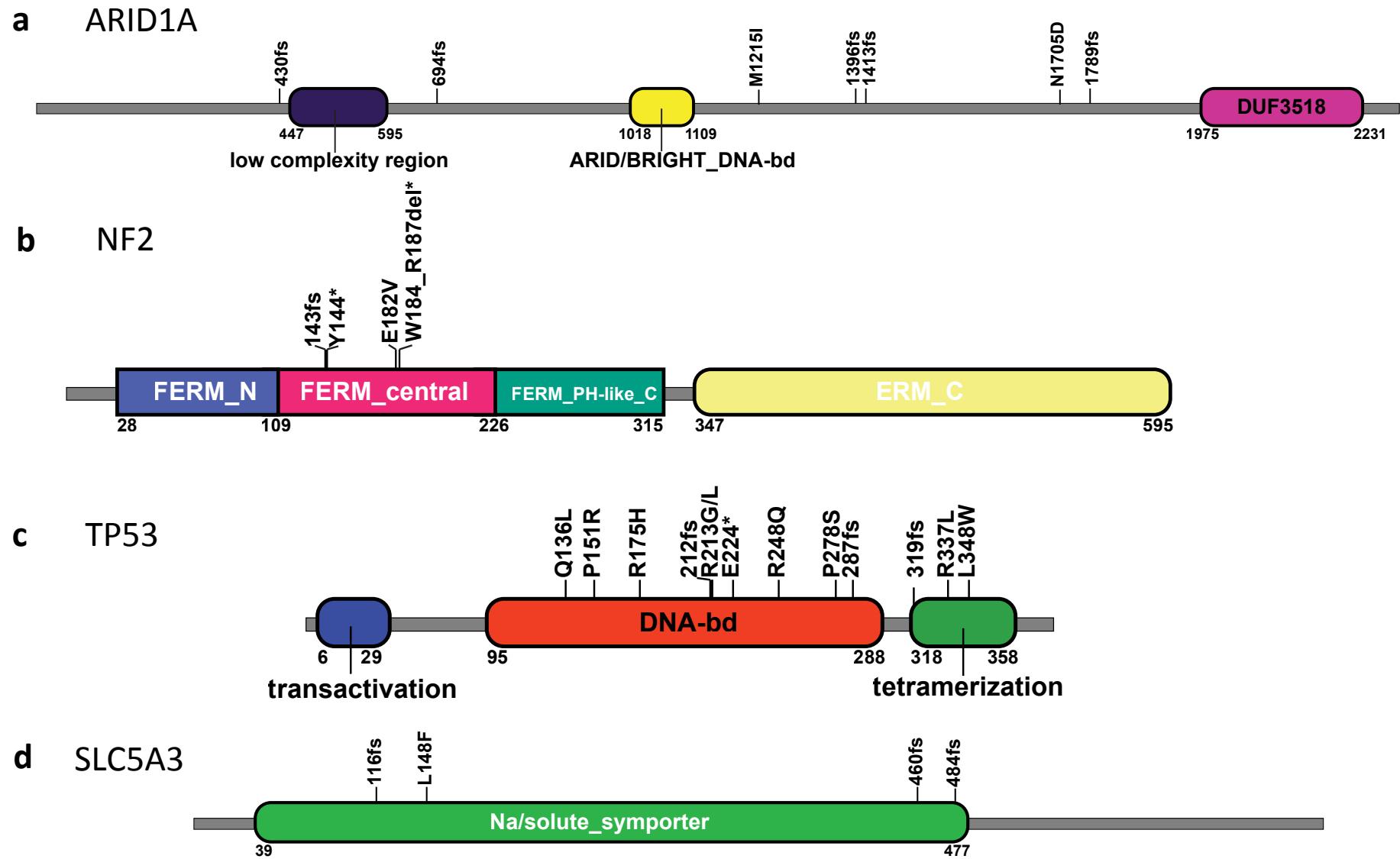
**Supplementary Fig. 5.** (a) Frequency of each mutation-type (C>A, C>G, C>T, T>A, T>C, T>G) in the context of 5' and 3' flanking bases in cancers as indicated. (b) Unsupervised hierarchical clustering of the mutation types at the sample level in pRCC, chRCC and RO. (c) Clustering by mutation signatures at the cancer type levels show contribution of each signature to a particular cancer type. Abbreviations used - thca – thyroid carcinoma, ov - ovarian serous cystadenocarcinoma, hnsc – head and neck squamous cell carcinoma, lusc – lung squamous cell carcinoma, luad – lung adenocarcinoma, gbm - glioblastoma multiforme, skcm – skin cutaneous melanoma).



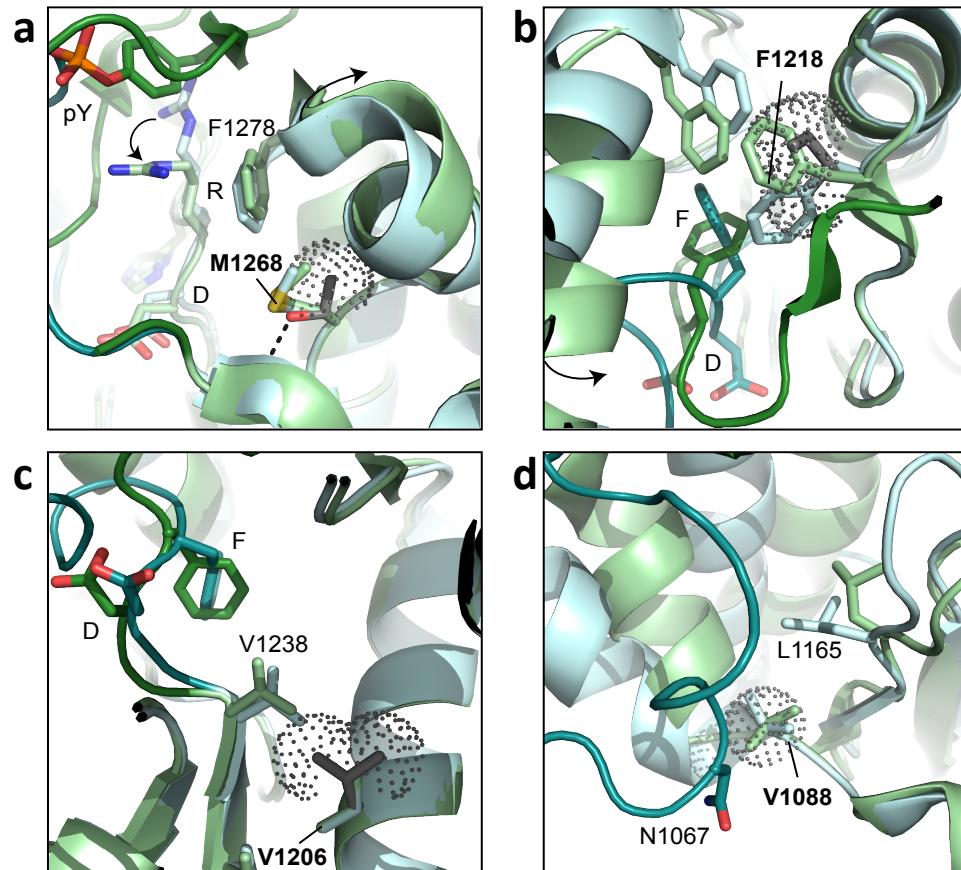
**Supplementary Fig. 6.** Plot of percentage of germline and somatic mutations scored as deleterious by both Polyphen and SIFT.



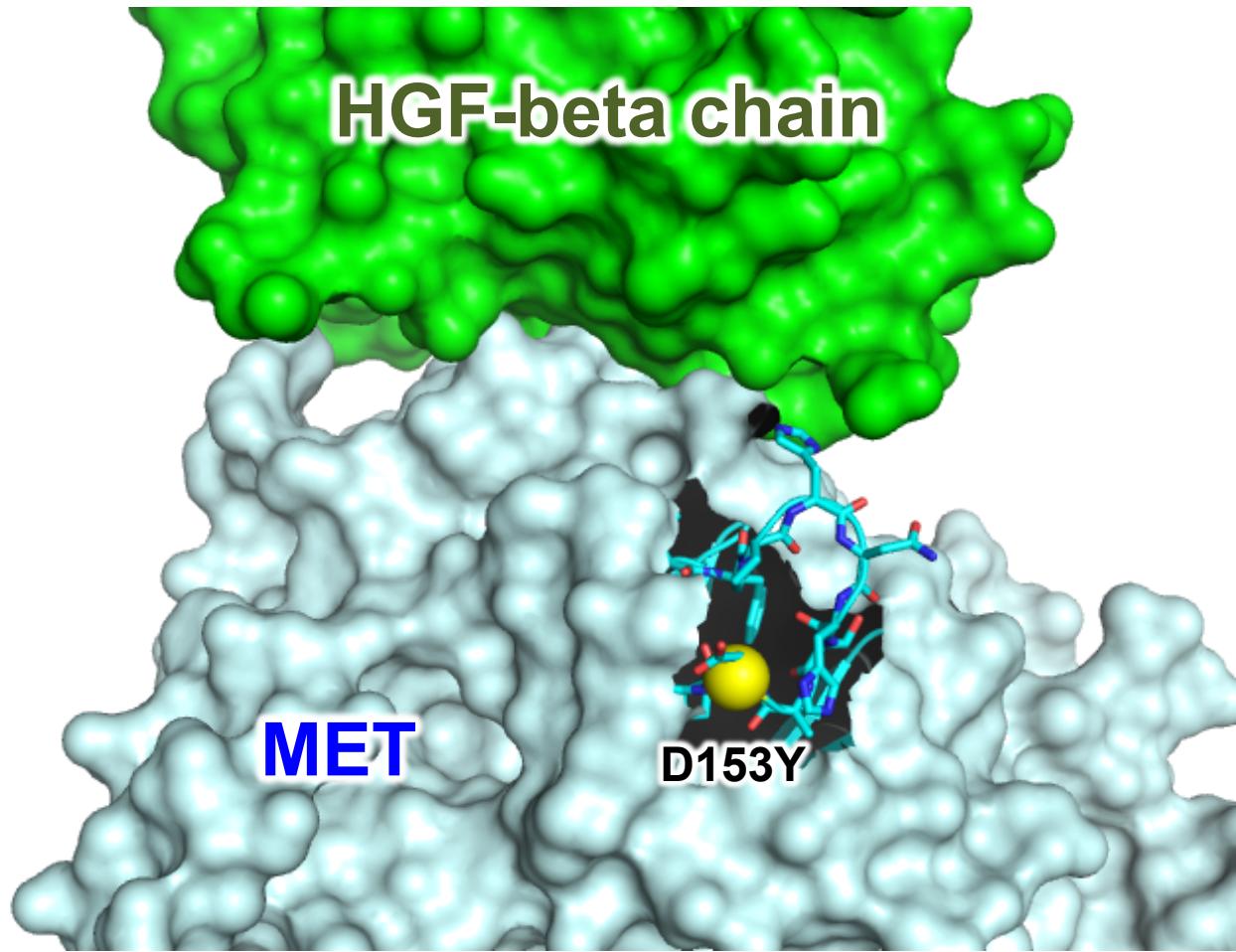
**Supplementary Fig. 7.** Distribution of frequency of deleterious mutations in one thousand Monte Carlo simulations. Black asterisk shows frequency of non-synonymous deleterious mutations observed in the exome data for each of the subtypes as indicated. P-values calculated as proportion of simulations that had a higher frequency than the observed frequency is shown. (TCGA - Nature 2013, 46:860). Boxplot defined in **Fig.1** legend in the main text.



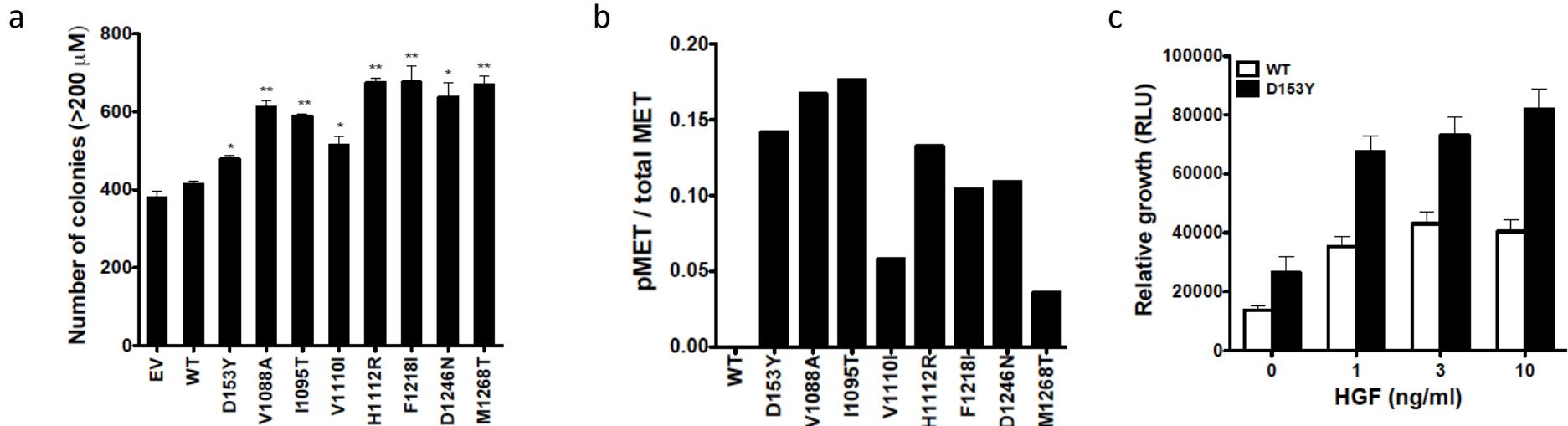
**Supplementary Fig. 8.** Somatic mutations in select genes shown on a cartoon of the protein along with the domains. \* = stop. ARID1A: ARID/BRIGHT DNA-binding domain; DUF3518: Domain of unknown function (pfam database) FREM: F = 4.1 protein, E = ezrin, R = radixin and M = moesin, PH: Pleckstrin homology; ERM (pfam - PF00769): Ezrin/radixin/moesin family; DNA-bd: DNA-binding domain & SLC5A3: Sodium:solute symporter family.



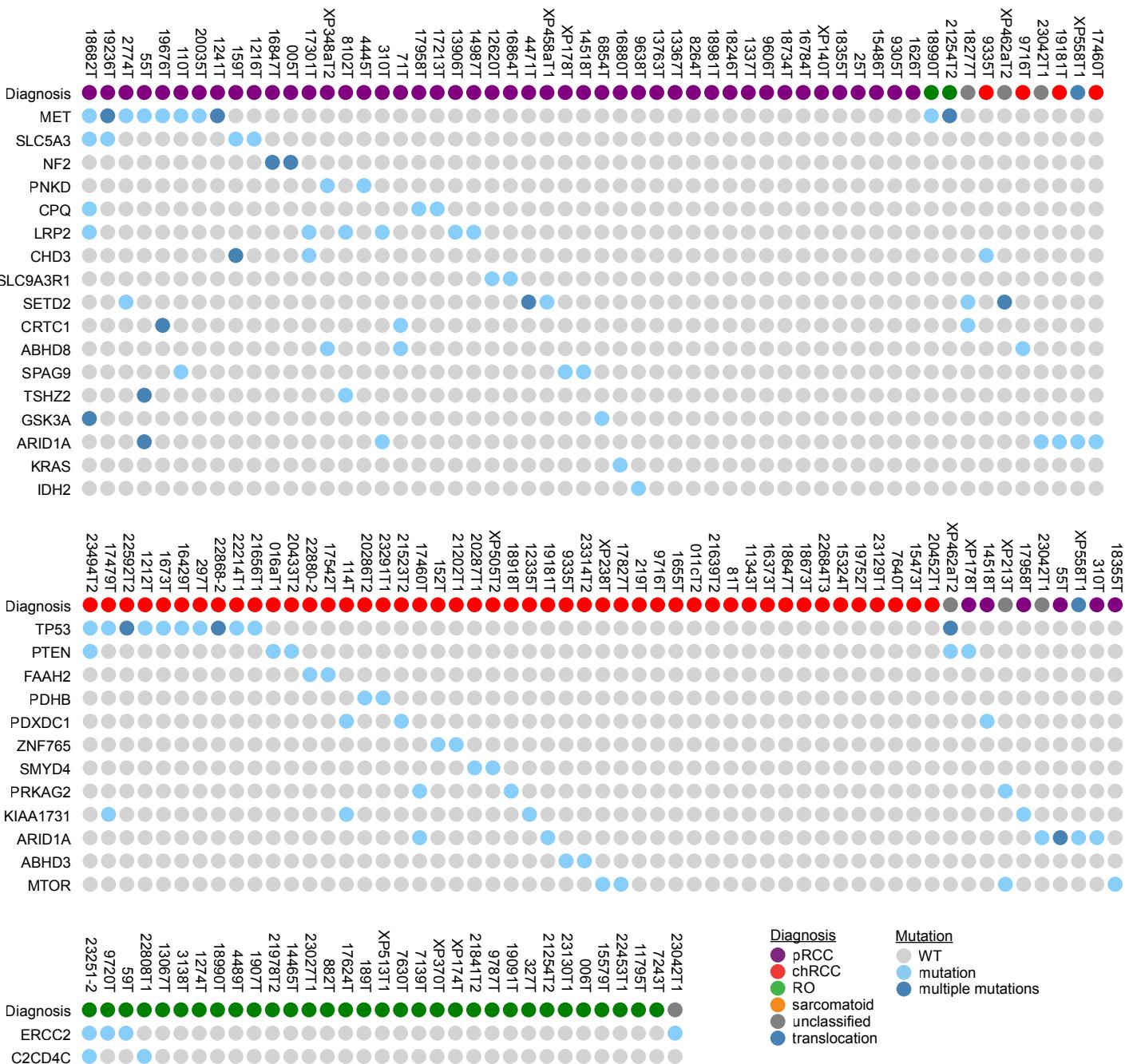
**Supplementary Fig. 9.** MET mutations favor the activated state of the kinase. MET tyrosine kinase domain mutations (labeled in bold with wild type residue depicted in stick and mutated alterations depicted in gray dots) were superimposed and compared in the inactive (PDB: 2g15, colored cyan) and active (PDB: 3r70, colored green) states. Both the activation loop and the N-terminal loop are significantly altered upon activation and are colored in darker hues. (a) M1268 positions the side chain of F1278 (numbered F1260 in the PDBs) adjacent to the catalytic loop HRD motif. Upon activation the HRD motif R side chain flips towards a phosphorylated Y (pY). This flip results in subtle movements of residues surrounding M1268: the F1278 side chain moves away from the catalytic loop and the connecting helix rotates away from the M1268. Mutation to T shifts the bulk of the side chain mass towards the C-alpha and introduces a polar group positioned to form a hydrogen bond (black dashed line) with a backbone oxygen. The T side chain should favor the helical rotation toward the activated state. (b) F1218 is adjacent to the activation loop DFG motif and the N-lobe helix. F1218 is in the C-terminal lobe of the kinase domain, interacting with residue F1241 of the DFG-motif (Fig. 2e). In the “resting” state, the A-loop and DFG of MET adopts an autoinhibited conformation (*Proc. Natl. Acad. Sci. USA* 103, 3563-8 (2006)). The side chain flips away from the DFG motif F in the activated state to position the N-lobe helix. Mutation of F1218 to I shifts the bulk of the side chain mass to mimic the flipped F and should favor the activated state N-lobe helical rotation. (c) V1206 is adjacent to a residue (V1088) positioned just prior to the activation loop. Mutation to L would increase the side chain bulk, which might be better tolerated in the activated state. Two previously identified mutations in this general area, D1246N and M1268T, were found to reduce the activation threshold, thus leading to hyperactivity (*Proteins* 44, 32-43 (2001)). Mutation F1218I potentially exerts similar activating effects. (d) V1088 marks the start of an N-terminal helix that rotates upon activation. The V1088 side chain positions the rotating helix, the N-terminal loop (through interaction with N1067), and the N-lobe (through interaction with L1165) in the inactive state. Mutation to A would remove the interactions with N1067 and L1165 and allow the helix to rotate to the activated state. Further, V1088 and I1095 are proximity to the juxtamembrane domain, which has been shown to regulate kinase activity (*J. Biol. Chem.* 269, 1815-20 (1994)) and hence it is likely that these mutations alter MET kinase activity.



**Supplementary Fig. 10.** MET extracellular mutation mapped on to MET kinase domain (PDB: 1ROP) and HGF β-chain (PDB: 1SHY).



**Supplementary Fig. 11. (a)** Quantitation of number of colonies formed by NIH3T3 cells expressing MET mutants shown in Fig. 2f. Data shown are mean ± SEM of duplicate experiments (\* $p<0.05$  and \*\* $p<0.01$ ). **(b)** pMET level in cells stably expressing MET WT and mutants shown in Fig. 2g. The intensity of the band in the western blot in Fig. 2g was measured using Image Studio Software (LI-COR Biotechnology, NE) **(c)**. MET mutant D153Y showed increased cell growth in response to HGF compare to MET WT. NIH3T3 cells stably expressing MET WT or D153Y mutant were treated with indicated HGF concentrations and cell growth was measured 3 days post treatment. Data shown mean ± SEM (six replicates).



**Supplementary Fig. 12.** Quilt plot depicting mutations in key genes in each nccRCC subtype.

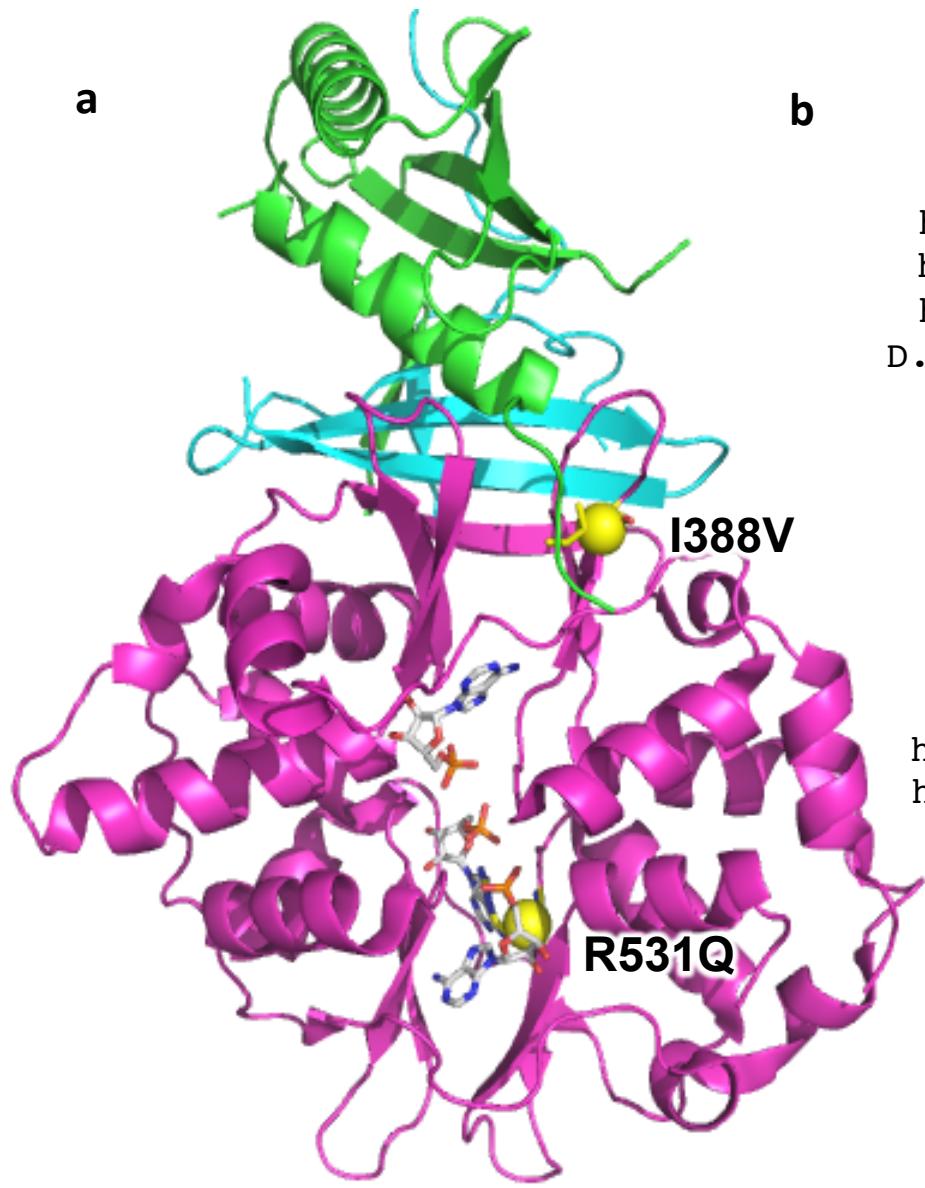
**a**

PDHB

**b**

<i>H. sapiens</i> <a href="#">NP_000916.2</a>	75	KKYGDKRIIDTPISEMGFAGIAVGAAMAGL	<b>R</b> PICEFMTNFSMQAIDQVI	124
<i>P. trloglodytes</i> <a href="#">NP_001233347.1</a>	75	KKYGDKRIIDTPISEMGFAGIAVGAAMAGL	<b>R</b> PICEFMTNFSMQAIDQVI	124
<i>B. Taurus</i> <a href="#">NP_001030512.2</a>	75	KKYGDKRIIDTPISEMGFAGIAVGAAMAGL	<b>R</b> PICEFMTNFSMQAIDQVI	124
<i>M. musculus</i> <a href="#">NP_077183.1</a>	75	KKYGDKRIIDTPISEMGFAGIAVGAAMAGL	<b>R</b> PICEFMTNFSMQAIDQVI	124
<i>R. novegicus</i> <a href="#">NP_001007621.1</a>	75	KKYGDKRIIDTPISEMGFAGIAVGAAMAGL	<b>R</b> PICEFMTNFSMQAIDQVI	124
<i>G. gallus</i> <a href="#">NP_001185549.1</a>	75	KKYGDKRIIDTPISEMGFTGIAVGAAMAGL	<b>R</b> PVCEFMTNFSMQAIDQVI	124
<i>X. tropicalis</i> <a href="#">NP_998319.1</a>	75	KKYGDKRIIDTPITEMGFAGIAVGAAMAGL	<b>R</b> PICEFMTNFSMQAIDQVI	124
<i>D. rerio</i> <a href="#">NP_651668.1</a>	71	KKYGDKRVIDTPITEMGFAGIAVGAAMAGL	<b>R</b> PVCEFMTWNFSMQAIDHII	120
<i>D. melanogaster</i> <a href="#">XP_311527.2</a>	70	KKYGDKRVIDTPITEMGFAGIAVGAAMAGL	<b>R</b> PVCEFMTNFSMQAIDHVI	119
<i>A. gambiae</i> <a href="#">NP_500340.1</a>	67	KKHGDKRVVDTPITEMGFAGIAVGAAGAGL	<b>R</b> PICEFMTNFSMQAIDQII	116

**Supplementary Fig. 13. (a)** Somatic mutations in PDHB shown on a cartoon of the protein along with the domains. **(b)** Multiple sequence alignment showing the conservation of the mutated residue.



**V387S = constitutive activation**

I388V

human AMPK- $\gamma$ 1:	LFDAVSSLIRNKIHRLPV <b>IDPE</b>
human AMPK- $\gamma$ 2:	LFDAVYSLIKNKIHRLPV <b>IDPI</b>
human AMPK- $\gamma$ 3:	LFEAVYTLIKNRIHRLPV <b>LDPV</b>
D. melanogaster:	LYDAIKILIHRSRIHRLPV <b>IDPA</b>
S. cerevisiae:	LFEACLKMLESRSGR <b>IPIDQD</b>
C. elegans:	LLDAACILAEHRVHR <b>IPIDPL</b>
A. thaliana:	SVLECMEVFSKG <b>IHLALVPVES</b>
G. lamblia:	LFDTLRLLRSHSVHRLPI <b>IDDG</b>

**c**

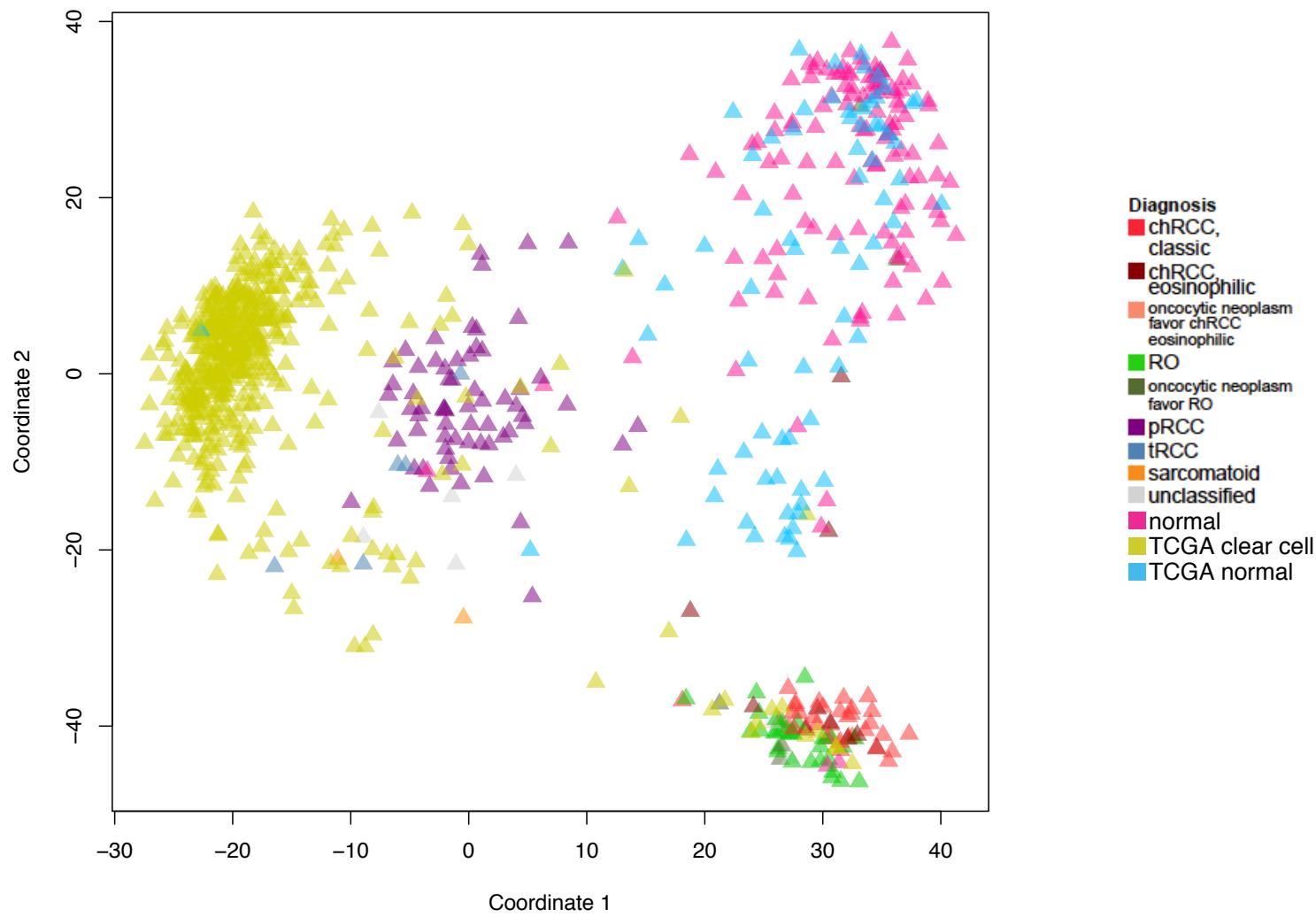
R299

human AMPK- $\gamma$ 1:	LETIINRLVEAEVHRLVVVDEND
human AMPK- $\gamma$ 2:	LETIVDRIVRAEVHRLVVVNEAD

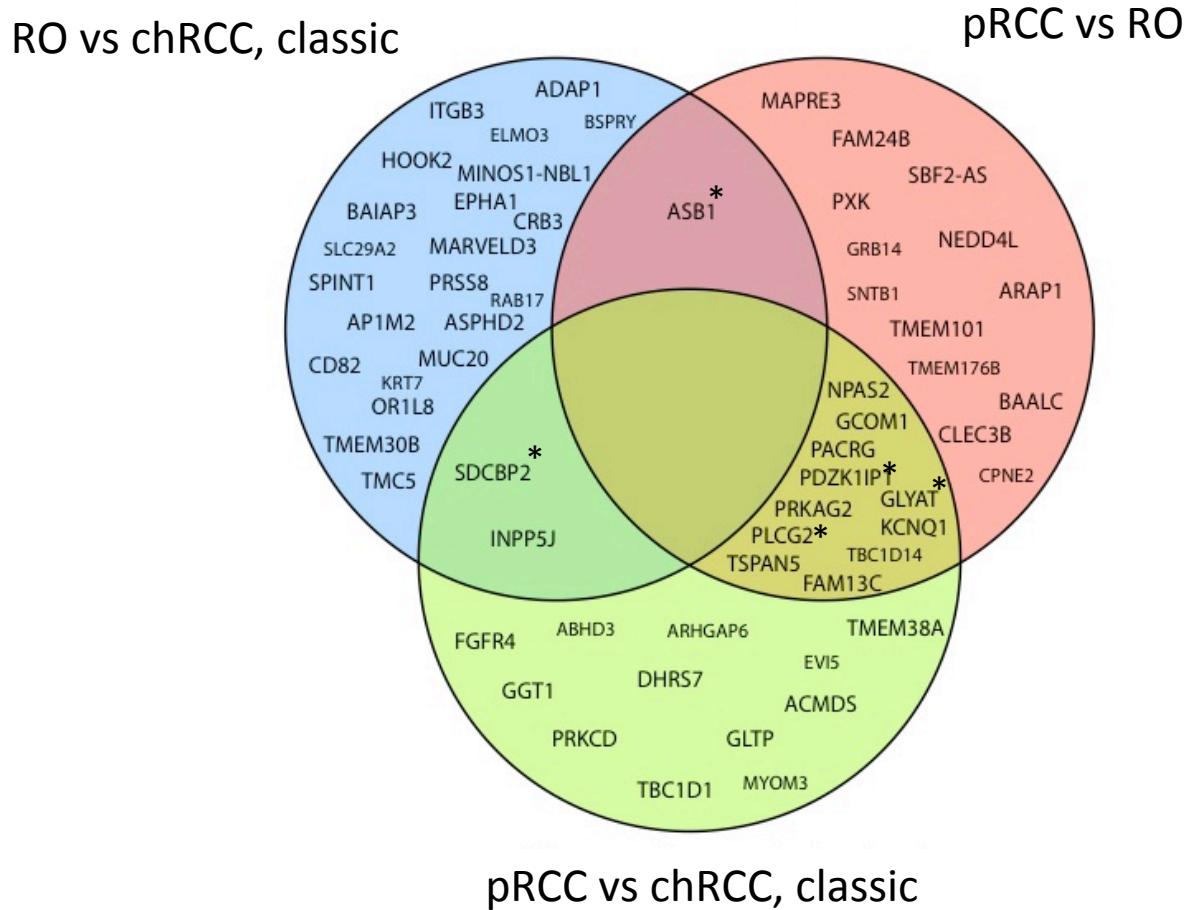
R531Q = constitutive activation

R531

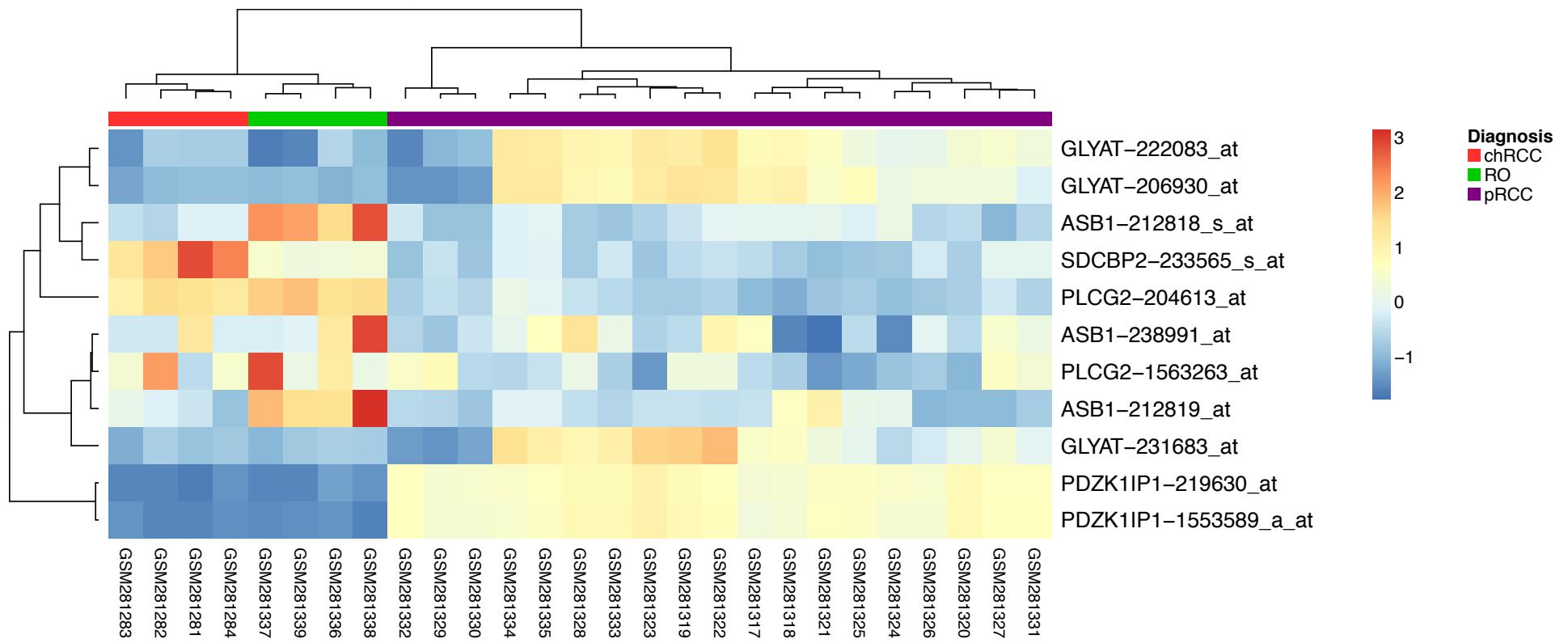
**Supplementary Fig. 14.** (a) PRKAG2 mutations mapped on to the structure (PDB: 2V8Q). (b) Alignment depicting the auto-inhibitory pseudosubstrate sequences in the second CBS motif of PRKAG2 (AMPK $\gamma$ ). The position of I388V mutant indicated in red and the position V387, mutation of which leads to constitutive activation (Scott, J.W. et. al. 2007) is indicated. (c) Alignment depicting the position of the recurrent mutation R531Q in PRKAG2 and its analogous position in PRKAG1 where a R299Q mutation was previously reported (Davis, C.F. et. al. 2014).



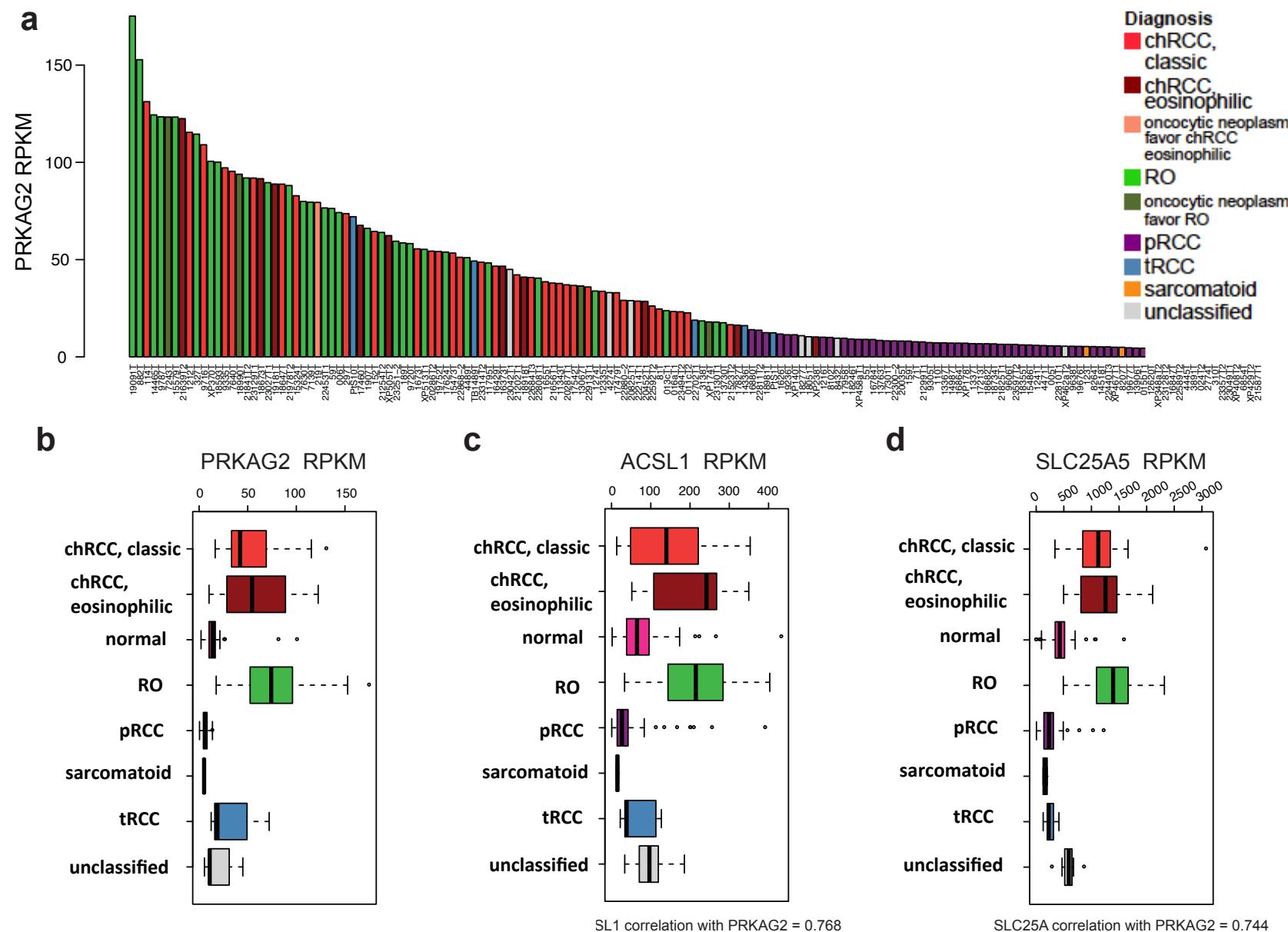
**Supplementary Fig. 15.** Multidimensional scaling plot based on the top 400 most variable genes in the variance stabilized expression matrix of the nccRCC data presented in this paper, combined with the published TCGA (ccRCC data -*Nature* 2013, 499:43-49) processed using the same methods show distinct sample clusters largely matching the sample diagnosis.



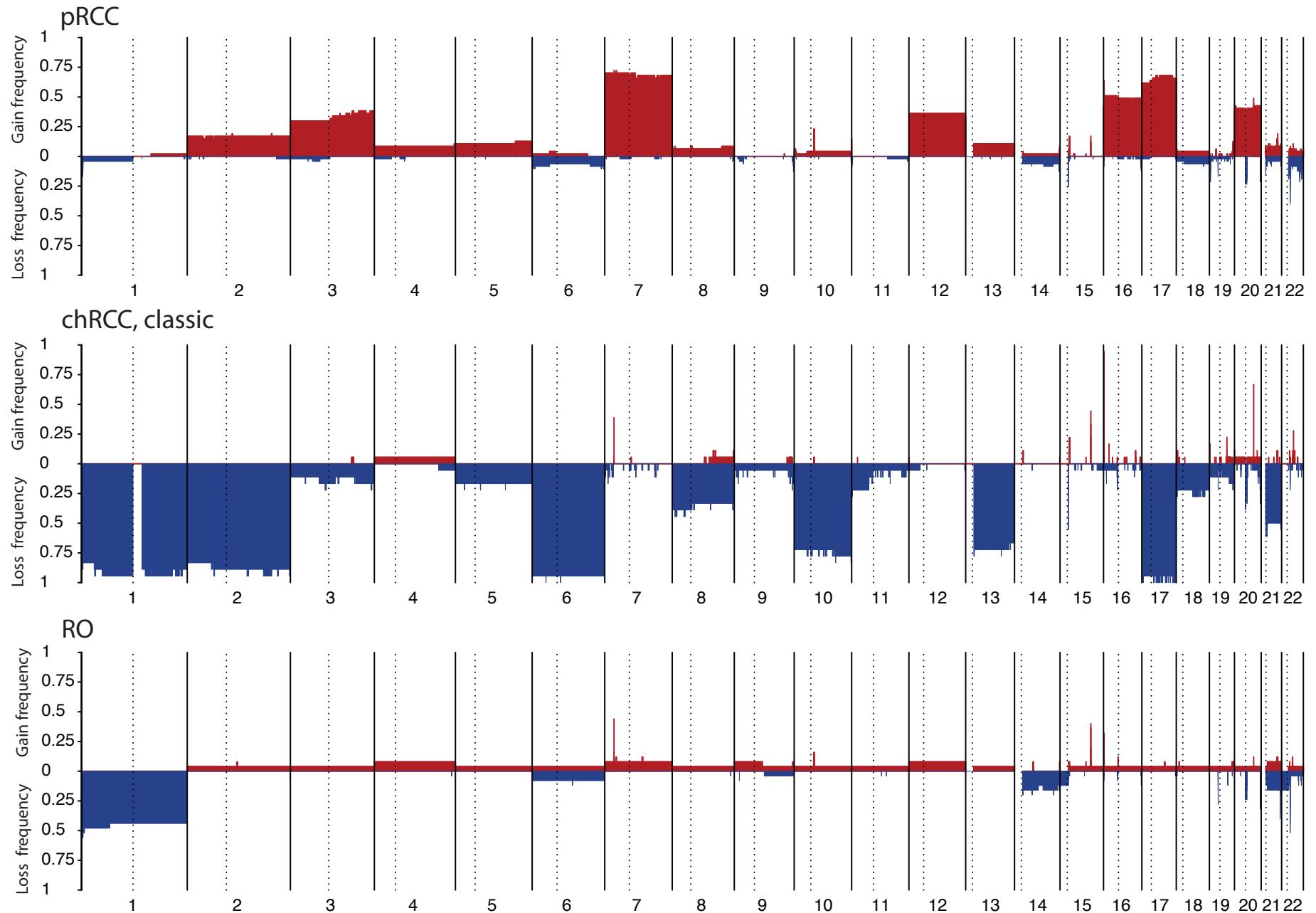
**Supplementary Fig. 16.** Differential gene expression was performed using edgeR with pairwise comparisons of discovery cohort chRCC classic, RO and pRCC samples (**Table 1** and **Supplementary Table 1**). The top 25 differentially expressed genes from each comparison are shown in the Venn diagram. Five genes\*, *ASB1*, *GLYAT*, *PDZK1IP1*, *PLCG2* and *SDCBP2*, derived from fourteen genes that were found differentially expressed in more than one of the two way comparisons were used for classifying nccRCC subtypes.



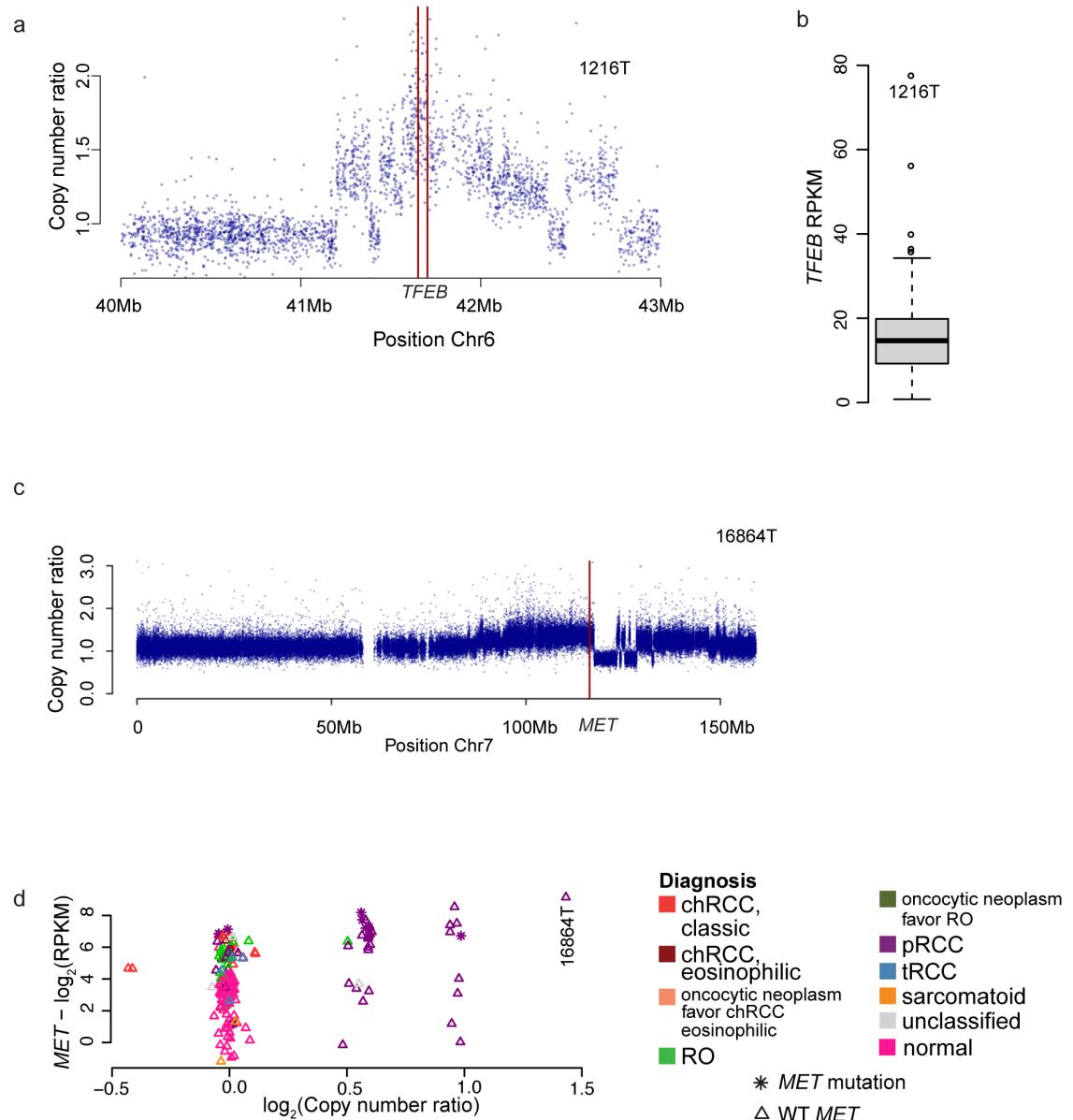
**Supplementary Fig. 17.** Unsupervised clustering of samples based on expression values obtained from GSE11151 (GEO database data set) for the probes corresponding to the indicated genes. The GSE11151 microarray data set used for validation consisted of 4 RO, 4 chRCC and 19 pRCC samples.



**Supplementary Fig. 18.** (a) Barplot of *PRKAG2* expression in tumors as measured by RNA-seq. (b) Boxplot of *PRKAG2* expression values grouped by subtype. (c) Boxplot of *ACSL1* expression values grouped by subtype. (d) Boxplot of *SLC25A5* expression values grouped by subtype. Boxplot defined in **Fig.1** legend in the main text.

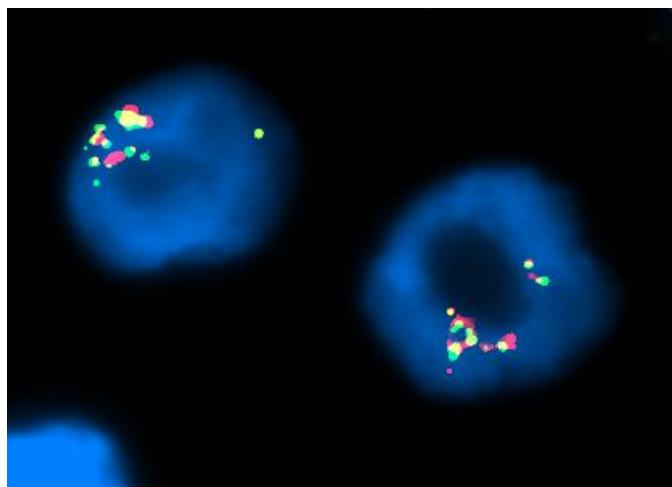


**Supplementary Fig. 19.** Summary of recurrent copy number alterations for each of the indicated nccRCC subtypes. In red are the frequencies of copy number gain events with a frequency of  $\log_2$  copy number ratios  $>0.45$ . In blue are the frequencies of copy number loss events with a frequency of  $\log_2$  copy number ratios  $<-0.45$ .

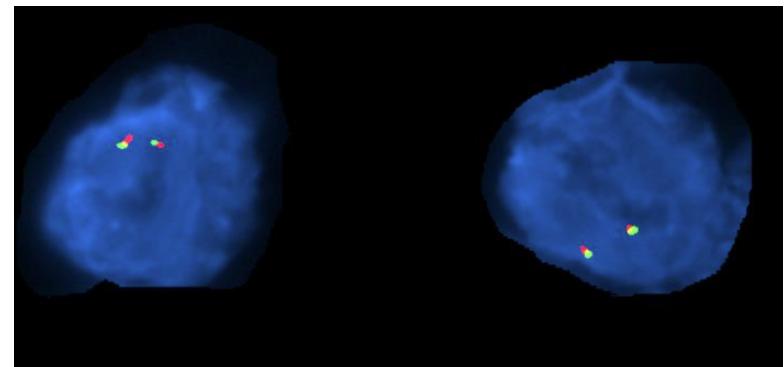


**Supplementary Fig. 20.** (a) Copy number ratio plot depicting a focal *TFEB* amplification in sample 1216T (b) Boxplot of *TFEB* expression in tumors show a high level of *TFEB* expression in sample 1216T (c) Copy number ratio plot depicting a 22Mb segmental amplification within chromosome 7 containing *MET* (d) Scatter plot of *MET* expression vs copy number. Boxplot (b) defined in **Fig.1** legend in the main text.

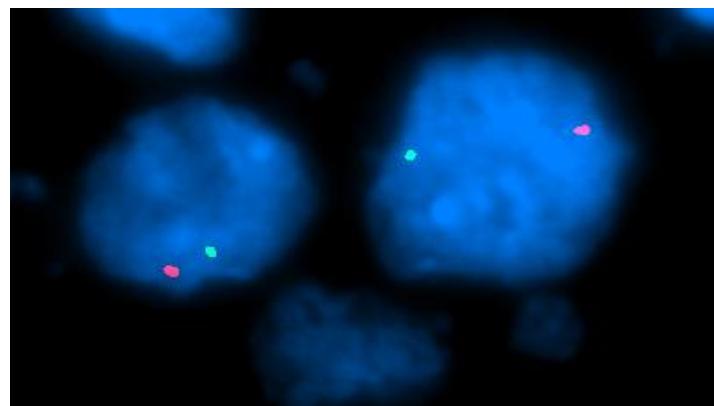
**a** **1216T1** TFEB amplification



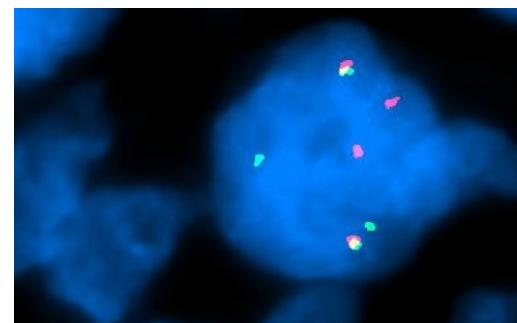
**b** Cancer Cell (normal TFE3)



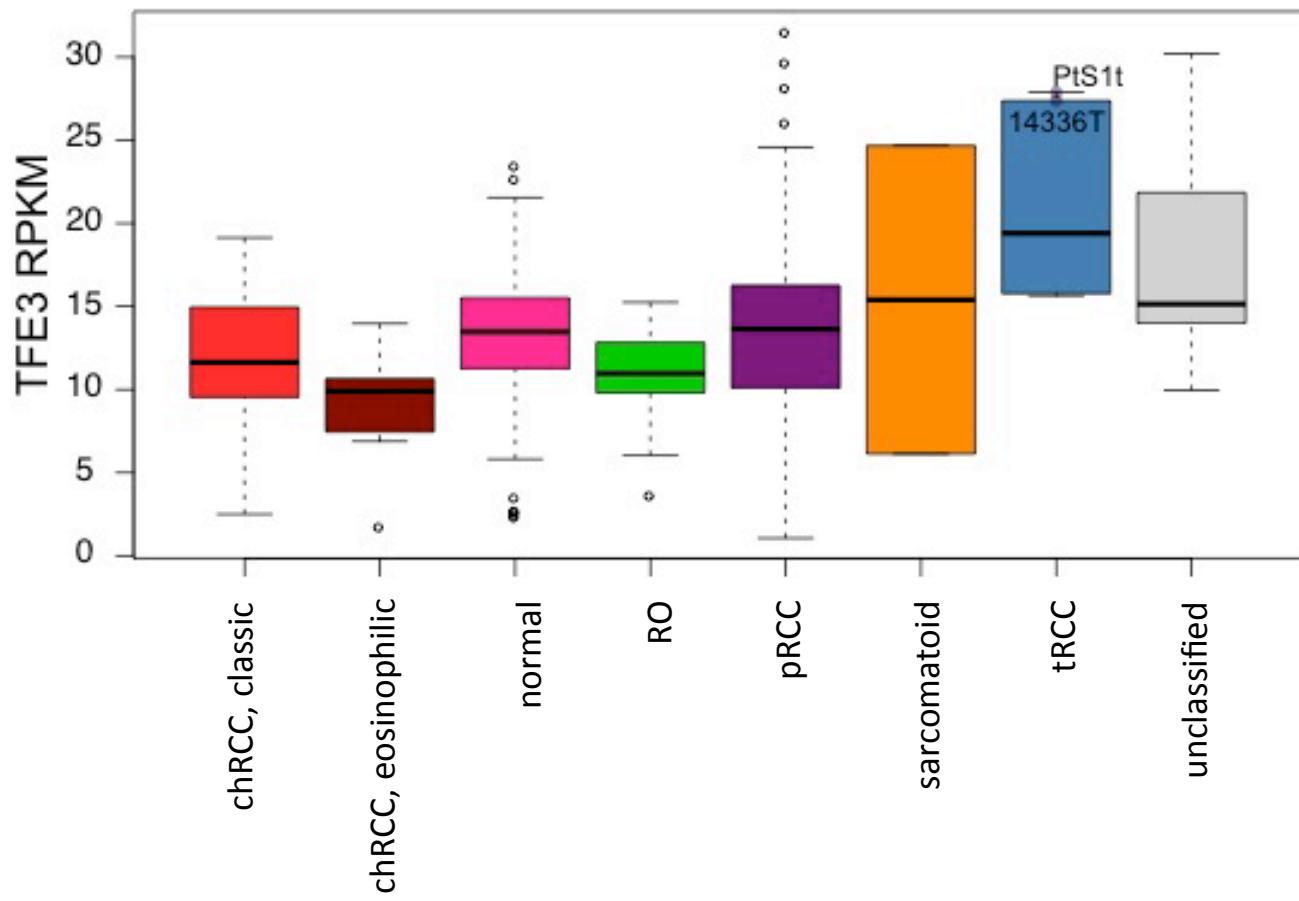
**c** **ptS10** (ASPSCR1-TFE3)



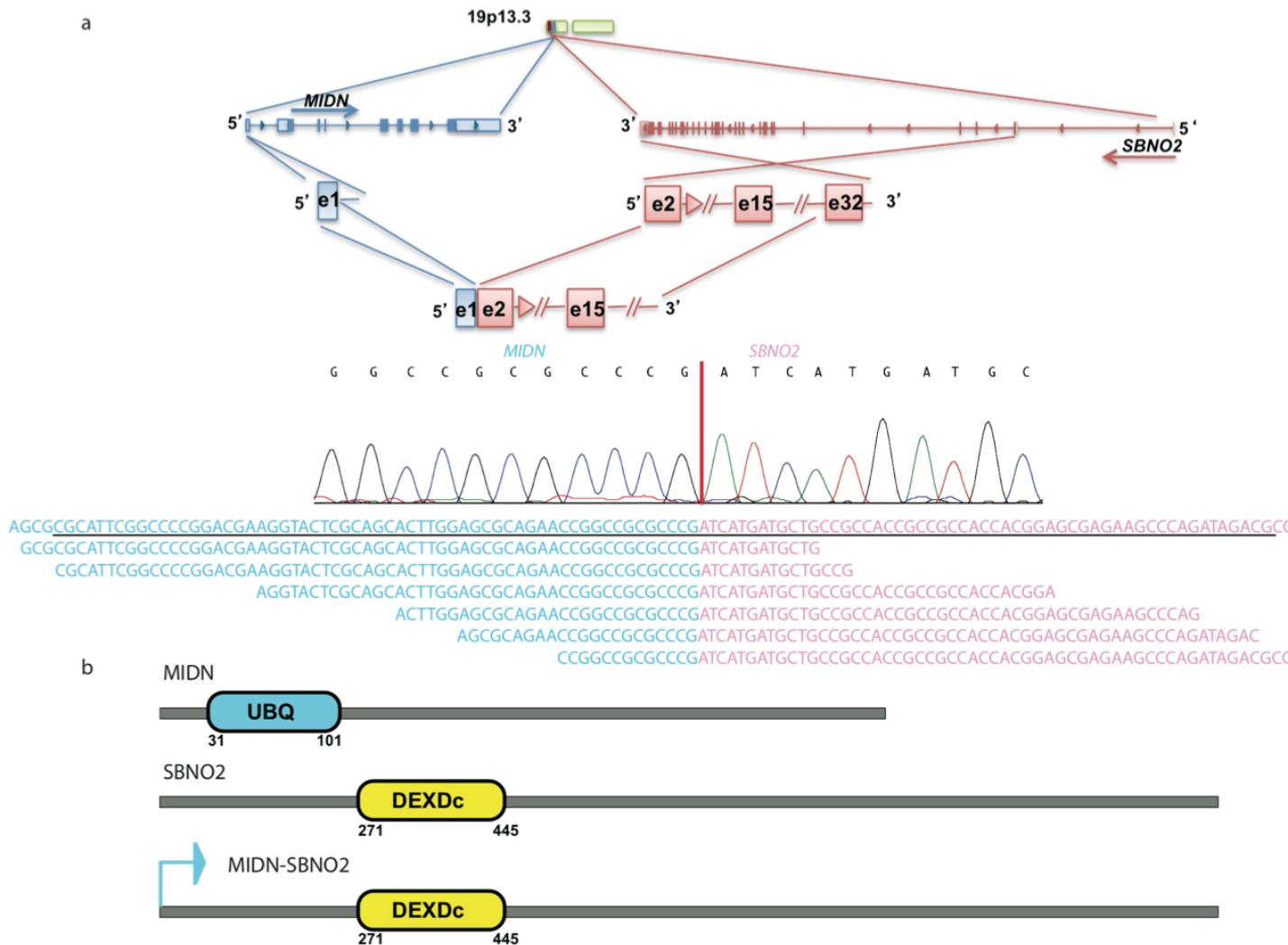
**d** **8432T1** (CLTC-TFEB)



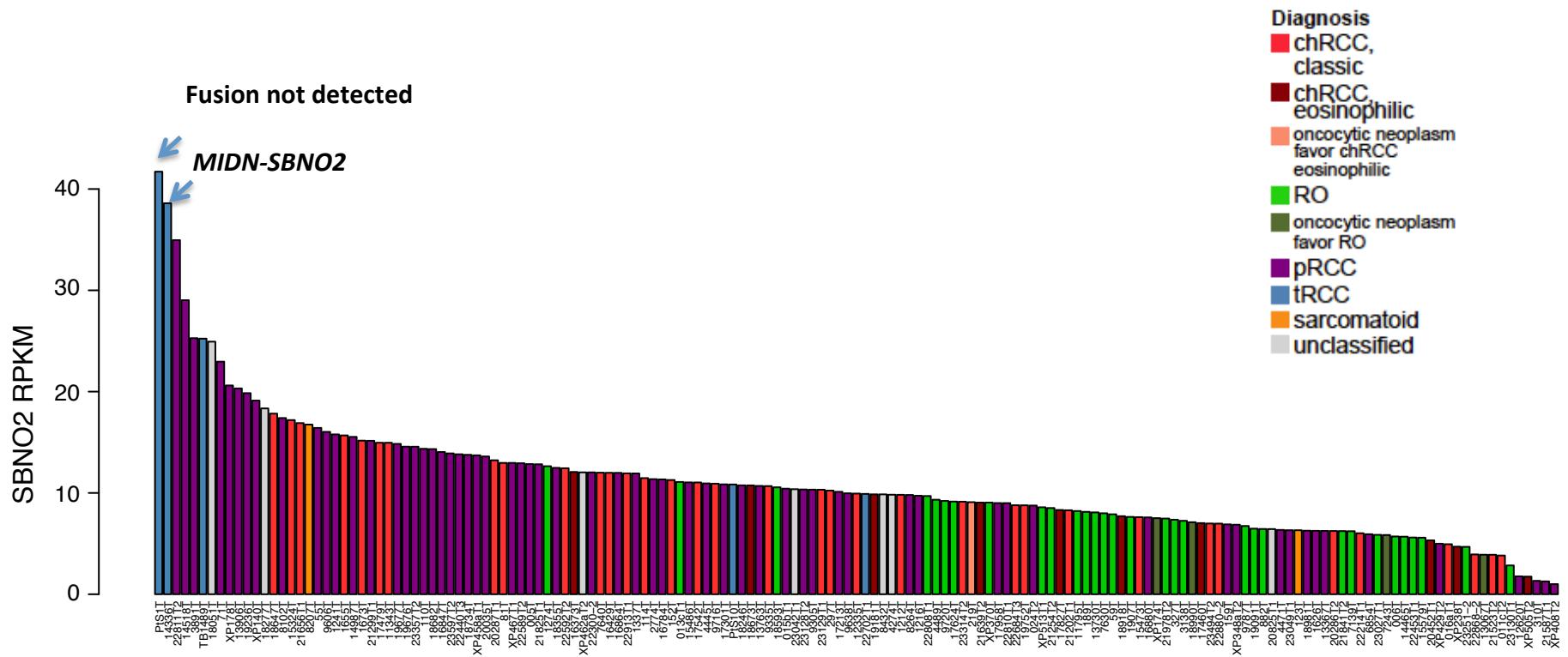
**Supplementary Fig. 21.** Fluorescence in situ hybridization (FISH) images on interphase nuclei of a DNA probe set (Agilent Technologies, CA) that hybridizes distal and proximal to the TFE3 and TFEB genes as indicated. **(a)** Increased adjacent red and yellow signal indicative of TFEB amplification, consistent with the TFEB amplification detected on the SNP array. **(b)** Tumor with two copies of adjacent red and green probes consistent with the presence of two unrecombined TFE3 alleles. **(c)** Separation of the green and red probes indicative of TFE3 translocation with loss of wild-type allele. **(d)** Two sets of separated green and red probes indicative of translocations involving two TFEB alleles. TFE3 probes used: TFE3 5' Xp11 labeled red and TFE3 3' Xp11 labeled green. TFEB probes used: TFEB 5' 6p21.1 labeled red and TFEB 3' 6p21.1 labeled green.



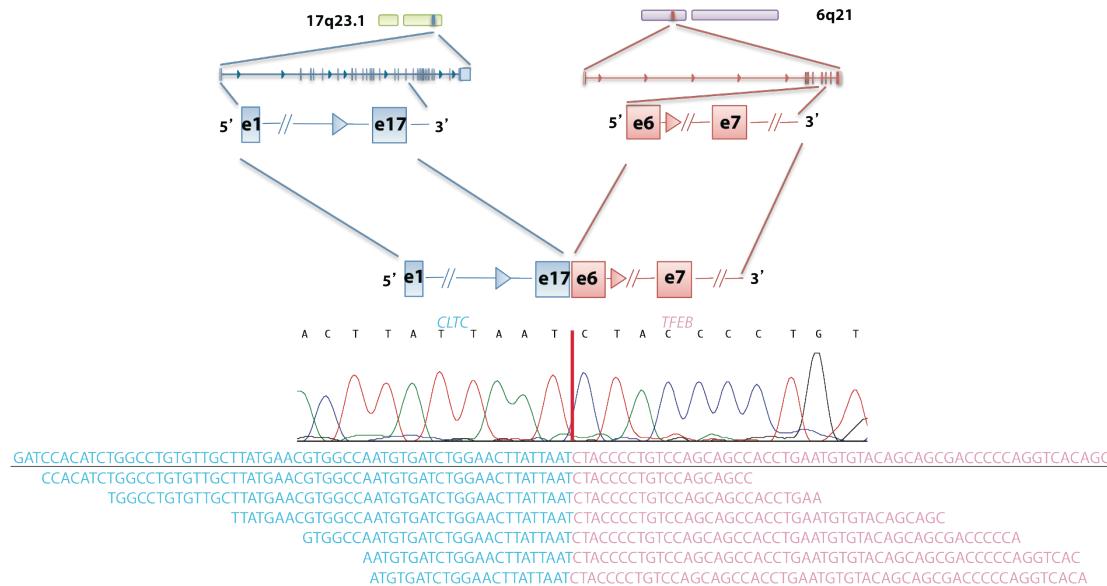
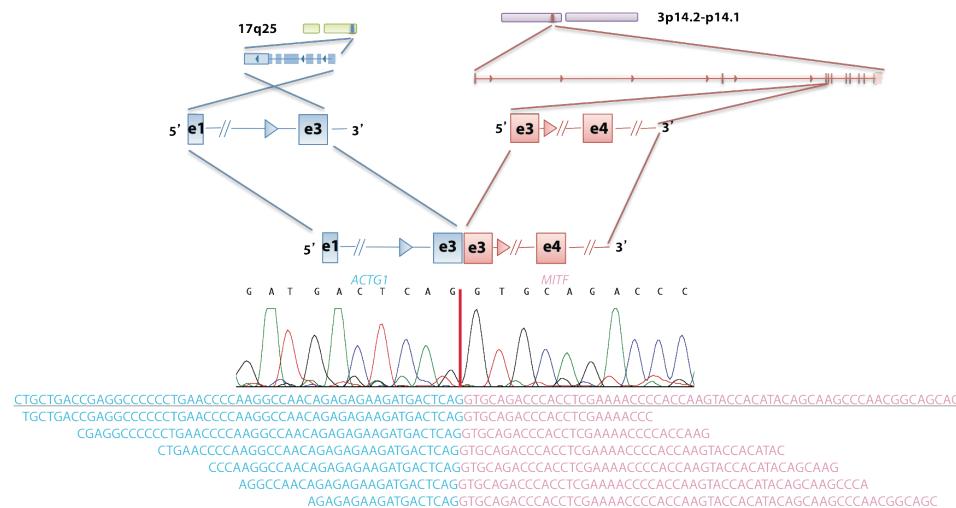
**Supplementary Fig. 22.** Boxplot of *TFE3* expression in the indicated nccRCC subtypes. Boxplot defined in **Fig.1** legend in the main text.



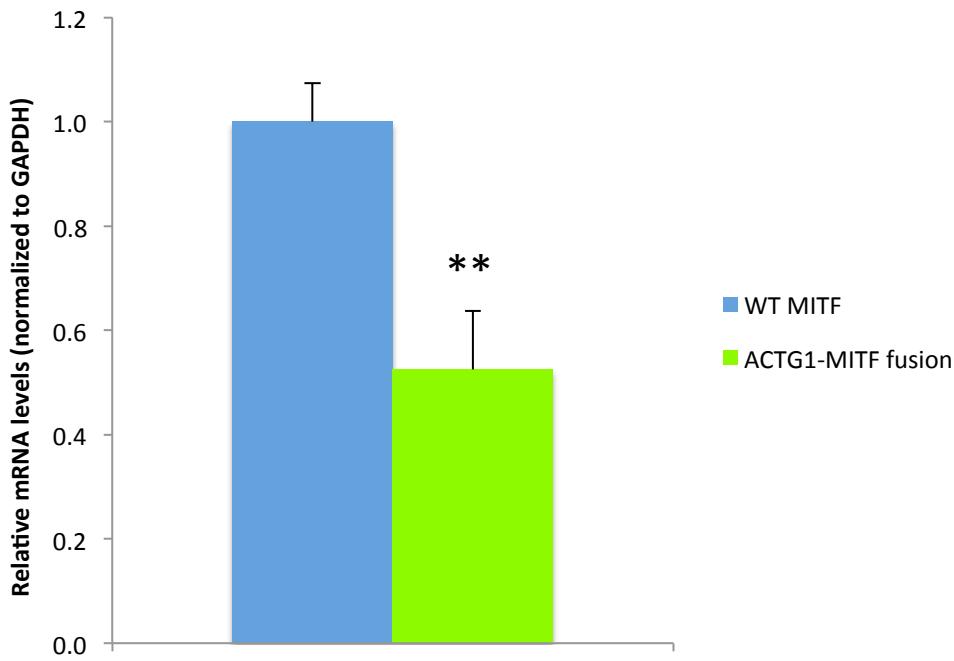
**Supplementary Fig. 23.** *MIDN-SBNO2* gene fusion **(a)** Cartoon depicting the location, orientation and exon-intron architecture of *MIDN-SBNO2* fusion on the genome. The read evidence for *MIDN*(e1)-*SBNO2*(e2) fusion identified using RNA-Seq data are shown. Representative Sanger sequencing chromatogram of the RT-PCR derived products confirming the *MIDN*(e1)-*SBNO2*(e2) fusion junction. **(b)** Schematic of the resulting *MIDN-SBNO2* fusion protein. Accession number for proteins used in the cartoon in panel b are: MIDN - NP\_796375.3 ; SBNO2 - NP\_055778.2



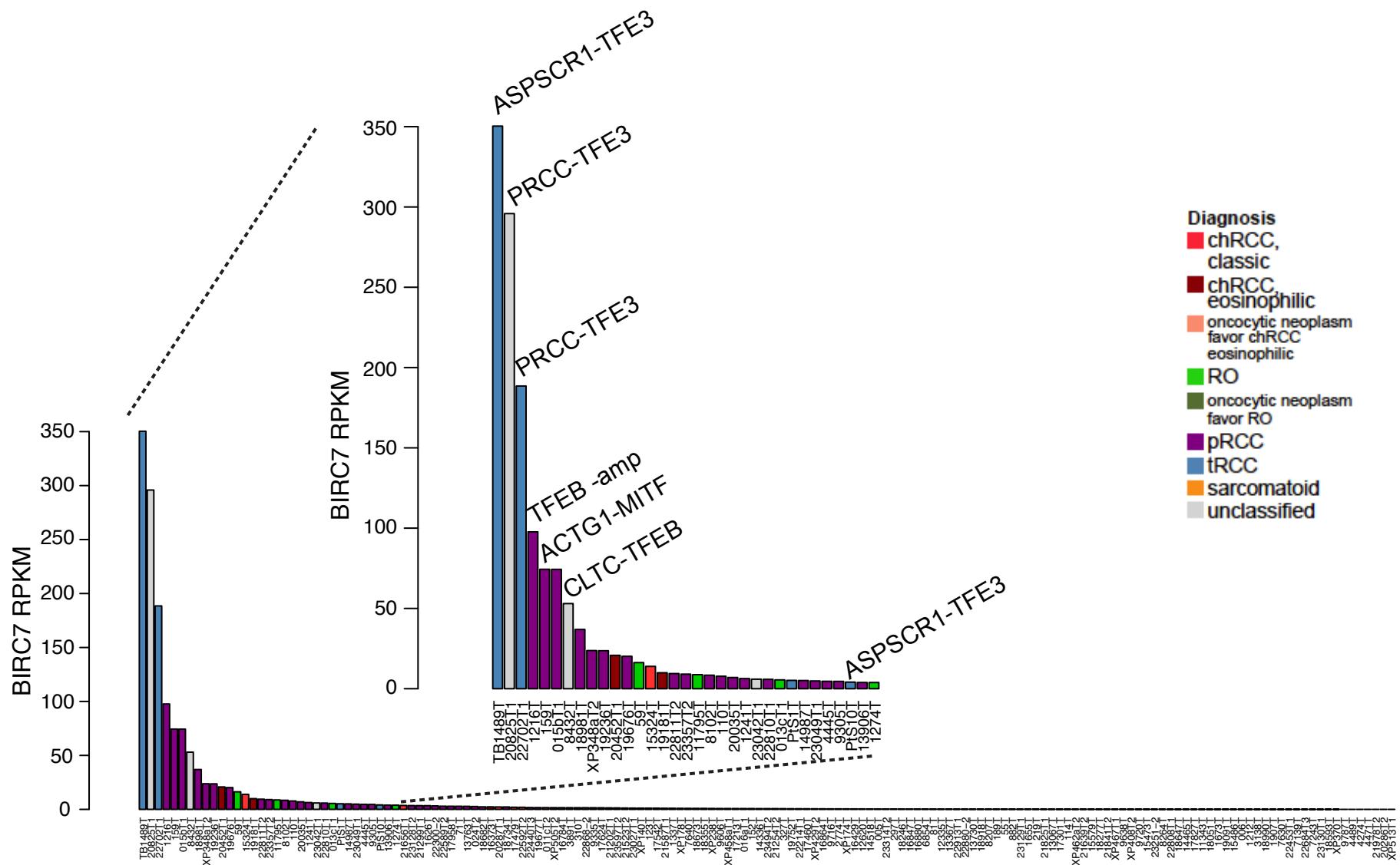
### **Supplementary Fig. 24.** Barplot of *SBNO2* expression in tumors as measured by RNA-seq.

**a****b**

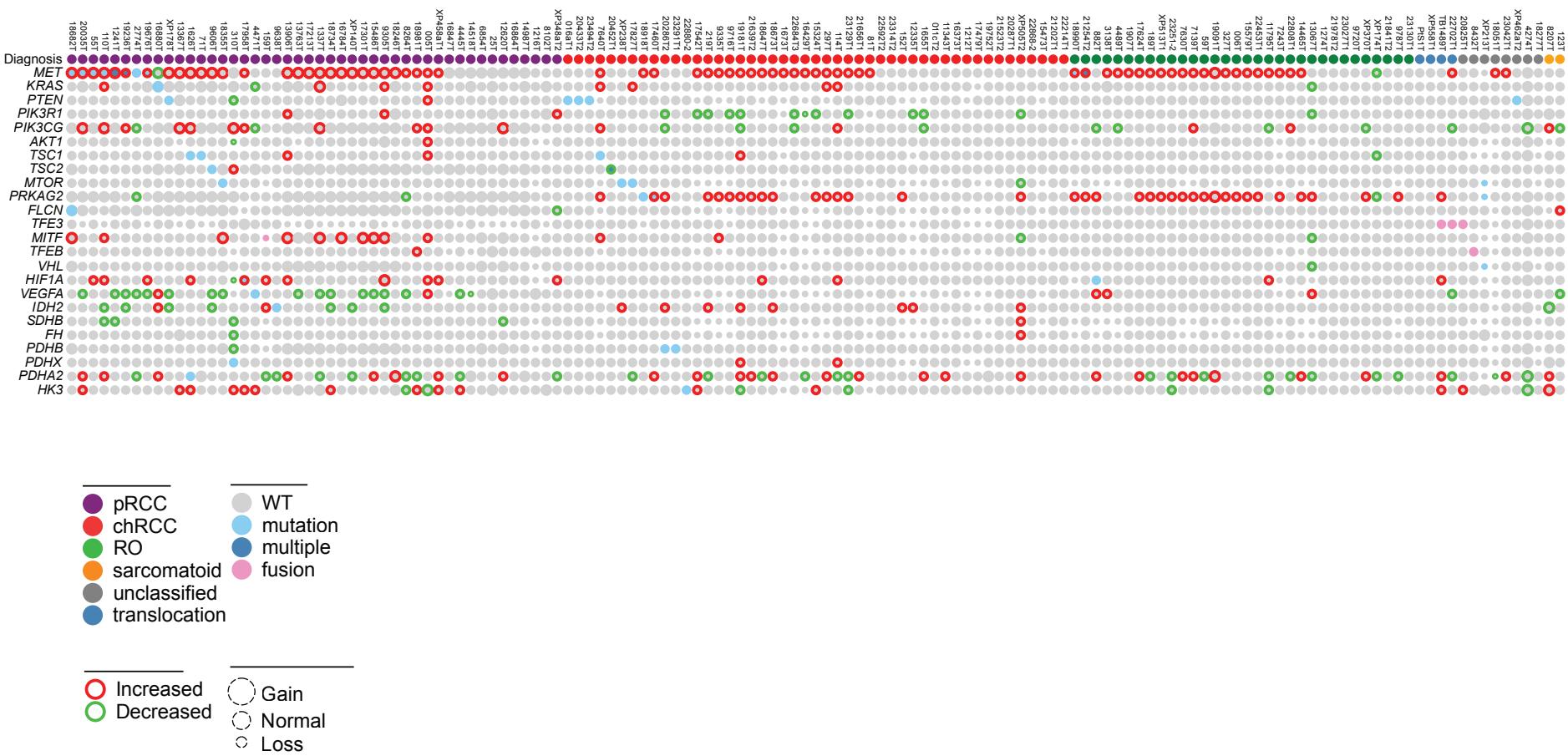
**Supplementary Fig. 25.** (a) Cartoon depicting the location, orientation and exon-intron architecture of *CLTC*-*TFEB* fusion on the genome. The read evidence for *CLTC*(e17)-*TFEB*(e6) fusion identified using RNA-seq data are shown. Representative Sanger sequencing chromatogram of the RT-PCR derived products confirming the *CLTC*(e17)-*TFEB*(e6) fusion junction. (b) Cartoon depicting the location, orientation, exon-intron architecture of *ACTG1*-*MITF* fusion on the genome, the read evidence for *ACTG1*(e3)-*MITF*(e3) fusion identified using RNA-seq data and a representative Sanger sequencing chromatogram of RT-PCR derived product confirming the *ACTG1*(e3)-*MITF*(e3) fusion junction are shown



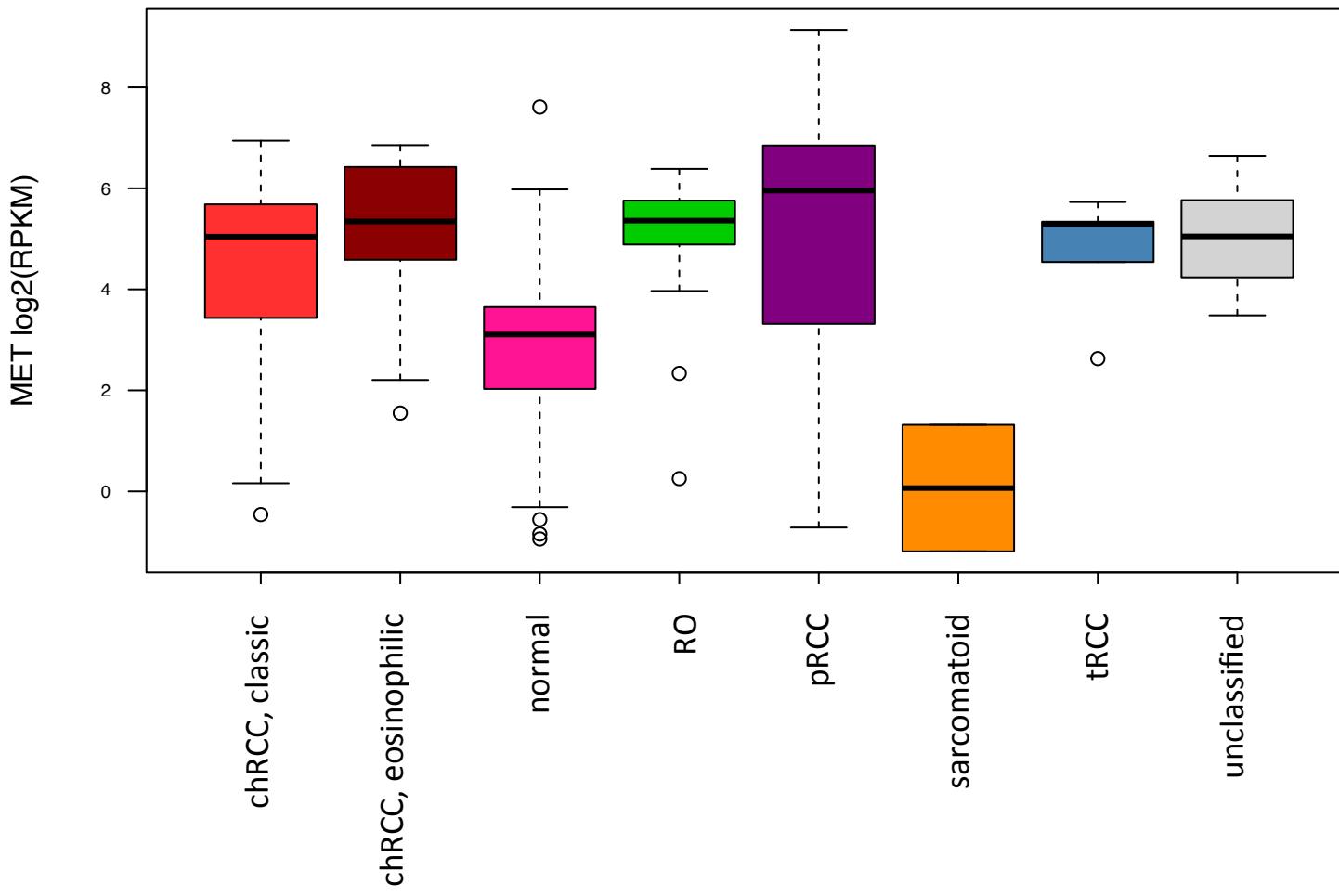
**Supplementary Fig. 26.** Expression levels of WT MITF and ACTG1-MITF in transiently transfected 293T cells. This was used to normalize the expression of target genes shown in Fig. 5a. The values shown are from three replicates. (error bar represents SEM; \*\* $p<0.01$ ).



**Supplementary Fig. 27.** Expression of *BIRC7* in tumors with MITF/TFE translocation or fusion compared to samples without a translocation. Expression of *BIRC7* was significant in samples with MITF/TFE event (*t*-test P=0.002308467).



**Supplementary Fig. 28.** Quilt plot depicting the mutation, expression, fusion and copy number changes across select genes involved in MET signaling and metabolism. Each column represents data from a single tumor sample. Genes highlighted as being over or under expressed showed a  $\log_2$  ratio of  $\geq 2$  and  $\leq -2$  respectively. Copy number gain events represent  $\log_2$  copy number ratios  $\geq 0.45$  and copy number loss events represent  $\log_2$  copy number ratios  $\leq -0.45$ .



**Supplementary Fig. 29.** Boxplot of *MET* expression in tumors by subtype as measured by RNA-seq. Increase in *MET* expression in tumors compared to normal observed in pRCC ( $P = 2.7213 \times 10^{-37}$ ,  $\log_2$  fold change = 2.82757), chRCC ( $P=1.280357 \times 10^{-17}$ ,  $\log_2$  fold change = 2.22951) and RO ( $P = 7.083757 \times 10^{-18}$ ,  $\log_2$  fold change = 2.105337). Boxplot defined in **Fig.1** legend in the main text.